

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

Основан в 1997 г.

Периодичность 8 выпусков в год

DOI 10.18699/vjgb-24-01

Учредители

Сибирское отделение Российской академии наук

Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук»

Межрегиональная общественная организация Вавиловское общество генетиков и селекционеров

Главный редактор

А.В. Кочетов – академик РАН, д-р биол. наук (Россия)

Заместители главного редактора

Н.А. Колчанов – академик РАН, д-р биол. наук, профессор (Россия)

И.Н. Леонова – д-р биол. наук (Россия)

Н.Б. Рубцов – д-р биол. наук, профессор (Россия)

В.К. Шумный – академик РАН, д-р биол. наук, профессор (Россия)

Ответственный секретарь

Г.В. Орлова – канд. биол. наук (Россия)

Редакционная коллегия

Е.Е. Андронов – канд. биол. наук (Россия)

Ю.С. Аульченко – д-р биол. наук (Россия)

О.С. Афанасенко – академик РАН, д-р биол. наук (Россия)

Д.А. Афонников – канд. биол. наук, доцент (Россия)

Л.И. Афтанас – академик РАН, д-р мед. наук (Россия)

Л.А. Беспалова – академик РАН, д-р с.-х. наук (Россия)

А. Бёрнер – д-р наук (Германия)

Н.П. Бондарь – канд. биол. наук (Россия)

С.А. Боринская – д-р биол. наук (Россия)

П.М. Бородин – д-р биол. наук, проф. (Россия)

А.В. Васильев – чл.-кор. РАН, д-р биол. наук (Россия)

М.И. Воевода – академик РАН, д-р мед. наук (Россия)

Т.А. Гавриленко – д-р биол. наук (Россия)

И. Гроссе – д-р наук, проф. (Германия)

Н.Е. Грунтенко – д-р биол. наук (Россия)

С.А. Демаков – д-р биол. наук (Россия)

И.К. Захаров – д-р биол. наук, проф. (Россия)

И.А. Захаров-Гезехус – чл.-кор. РАН, д-р биол. наук (Россия)

С.Г. Инге-Вечтомов – академик РАН, д-р биол. наук (Россия)

А.В. Кильчевский – чл.-кор. НАНБ, д-р биол. наук (Беларусь)

С.В. Костров – чл.-кор. РАН, д-р хим. наук (Россия)

А.М. Кудрявцев – чл.-кор. РАН, д-р биол. наук (Россия)

И.Н. Лаврик – д-р биол. наук (Германия)

Д.М. Ларкин – канд. биол. наук (Великобритания)

Ж. Ле Гуи – д-р наук (Франция)

И.Н. Лебедев – д-р биол. наук, проф. (Россия)

Л.А. Лутова – д-р биол. наук, проф. (Россия)

Б. Люгтенберг – д-р наук, проф. (Нидерланды)

В.Ю. Макеев – чл.-кор. РАН, д-р физ.-мат. наук (Россия)

В.И. Молодин – академик РАН, д-р ист. наук (Россия)

М.П. Мошкин – д-р биол. наук, проф. (Россия)

С.Р. Мурсалимов – канд. биол. наук (Россия)

Л.Ю. Новикова – д-р с.-х. наук (Россия)

Е.К. Потокина – д-р биол. наук (Россия)

В.П. Пузырев – академик РАН, д-р мед. наук (Россия)

Д.В. Пышный – чл.-кор. РАН, д-р хим. наук (Россия)

И.Б. Розозин – канд. биол. наук (США)

А.О. Рувинский – д-р биол. наук, проф. (Австралия)

Е.Ю. Рыкова – д-р биол. наук (Россия)

Е.А. Салина – д-р биол. наук, проф. (Россия)

В.А. Степанов – академик РАН, д-р биол. наук (Россия)

И.А. Тихонович – академик РАН, д-р биол. наук (Россия)

Е.К. Хлесткина – д-р биол. наук, проф. РАН (Россия)

Э.К. Хуснутдинова – д-р биол. наук, проф. (Россия)

М. Чен – д-р биол. наук (Китайская Народная Республика)

Ю.Н. Шавруков – д-р биол. наук (Австралия)

Р.И. Шейко – чл.-кор. НАНБ, д-р с.-х. наук (Беларусь)

С.В. Шестаков – академик РАН, д-р биол. наук (Россия)

Н.К. Янковский – академик РАН, д-р биол. наук (Россия)

Online edition

VAVILOV JOURNAL OF GENETICS AND BREEDING

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

Founded in 1997

Published 8 times annually

DOI 10.18699/vjgb-24-01

Founders

Siberian Branch of the Russian Academy of Sciences

Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences

The Vavilov Society of Geneticists and Breeders

Editor-in-Chief

A.V. Kochetov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

Deputy Editor-in-Chief

N.A. Kolchanov, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

I.N. Leonova, Dr. Sci. (Biology), Russia

N.B. Rubtsov, Professor, Dr. Sci. (Biology), Russia

V.K. Shumny, Full Member of the Russian Academy of Sciences, Dr. Sci. (Biology), Russia

Executive Secretary

G.V. Orlova, Cand. Sci. (Biology), Russia

Editorial board

O.S. Afanasenko, Full Member of the RAS, Dr. Sci. (Biology), Russia

D.A. Afonnikov, Associate Professor, Cand. Sci. (Biology), Russia

L.I. Aftanas, Full Member of the RAS, Dr. Sci. (Medicine), Russia

E.E. Andronov, Cand. Sci. (Biology), Russia

Yu.S. Aulchenko, Dr. Sci. (Biology), Russia

L.A. Bepalova, Full Member of the RAS, Dr. Sci. (Agricul.), Russia

N.P. Bondar, Cand. Sci. (Biology), Russia

S.A. Borinskaya, Dr. Sci. (Biology), Russia

P.M. Borodin, Professor, Dr. Sci. (Biology), Russia

A. Börner, Dr. Sci., Germany

M. Chen, Dr. Sci. (Biology), People's Republic of China

S.A. Demakov, Dr. Sci. (Biology), Russia

T.A. Gavrilenko, Dr. Sci. (Biology), Russia

I. Grosse, Professor, Dr. Sci., Germany

N.E. Gruntenko, Dr. Sci. (Biology), Russia

S.G. Inge-Vechtomov, Full Member of the RAS, Dr. Sci. (Biology), Russia

E.K. Khlestkina, Professor of the RAS, Dr. Sci. (Biology), Russia

E.K. Khusnutdinova, Professor, Dr. Sci. (Biology), Russia

A.V. Kilchevsky, Corr. Member of the NAS of Belarus, Dr. Sci. (Biology), Belarus

S.V. Kostrov, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia

A.M. Kudryavtsev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

D.M. Larkin, Cand. Sci. (Biology), Great Britain

I.N. Lavrik, Dr. Sci. (Biology), Germany

J. Le Gouis, Dr. Sci., France

I.N. Lebedev, Professor, Dr. Sci. (Biology), Russia

B. Lugtenberg, Professor, Dr. Sci., Netherlands

L.A. Lutova, Professor, Dr. Sci. (Biology), Russia

V.Yu. Makeev, Corr. Member of the RAS, Dr. Sci. (Physics and Mathem.), Russia

V.I. Molodin, Full Member of the RAS, Dr. Sci. (History), Russia

M.P. Moshkin, Professor, Dr. Sci. (Biology), Russia

S.R. Mursalimov, Cand. Sci. (Biology), Russia

L.Yu. Novikova, Dr. Sci. (Agricul.), Russia

E.K. Potokina, Dr. Sci. (Biology), Russia

V.P. Puzyrev, Full Member of the RAS, Dr. Sci. (Medicine), Russia

D.V. Pyshnyi, Corr. Member of the RAS, Dr. Sci. (Chemistry), Russia

I.B. Rogozin, Cand. Sci. (Biology), United States

A.O. Ruvinsky, Professor, Dr. Sci. (Biology), Australia

E.Y. Rykova, Dr. Sci. (Biology), Russia

E.A. Salina, Professor, Dr. Sci. (Biology), Russia

Y.N. Shavrukov, Dr. Sci. (Biology), Australia

R.I. Sheiko, Corr. Member of the NAS of Belarus, Dr. Sci. (Agricul.), Belarus

S.V. Shestakov, Full Member of the RAS, Dr. Sci. (Biology), Russia

V.A. Stepanov, Full Member of the RAS, Dr. Sci. (Biology), Russia

I.A. Tikhonovich, Full Member of the RAS, Dr. Sci. (Biology), Russia

A.V. Vasiliev, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

M.I. Voevoda, Full Member of the RAS, Dr. Sci. (Medicine), Russia

N.K. Yankovsky, Full Member of the RAS, Dr. Sci. (Biology), Russia

I.K. Zakharov, Professor, Dr. Sci. (Biology), Russia

I.A. Zakharov-Gezekhus, Corr. Member of the RAS, Dr. Sci. (Biology), Russia

Материалы конференции
«Устойчивость растений и микроорганизмов
к неблагоприятным факторам среды», Иркутск

- 5 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Репродуктивные взаимоотношения между представителями таксонов, морфологически близких к *Elymus caninus* (Poaceae: Triticeae). *Е.В. Шабанова, А.В. Агафонов, О.В. Дорогина*
- 15 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Регуляторные эффекты (p)ppGpp и индола на синтез цАМФ в клетках *Escherichia coli*. *Н.М. Кашеварова, Е.А. Хаова, А.Г. Ткаченко*
- 24 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Регуляторные эффекты полиаминов и индола на экспрессию факторов гипернатрии рибосом у *Escherichia coli* на уровне трансляции. *Е.А. Хаова, А.Г. Ткаченко*
- 33 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Экспрессия генов транспортеров ауксина в волокнах льна (*Linum usitatissimum*) при гравитответе. *Н.Н. Ибрагимова, Н.Е. Мокшина*
- 44 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Аскорбат-глутатионовый цикл в проростках пшеницы и риса при аноксии и последующей реаэрации. *В.В. Емельянов, Е.Г. Приказюк, В.В. Ласточкин, О.М. Арешева, Т.В. Чиркова*
- 55 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Морфологический и молекулярный анализ сортов роз из садовых групп грандифлора и розы Кордеса. *С.С. Юданова, О.В. Дорогина, О.Ю. Васильева*

Генетика растений

- 63 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Проявление селекционно-ценных признаков в потомстве мутанта сорго, несущего генетическую конструкцию для РНК-сайленсинга гена γ -кафирина. *Л.А. Эльконин, Н.В. Борисенко, Т.Е. Пылаев, О.А. Кенжегулов, С.Х. Сарсенова, Н.Ю. Селиванов, В.М. Панин*
- 74 **МЕТОДЫ И ПРОТОКОЛЫ**
Halo-RPD: в поисках мишеней РНК-связывающих белков растений. *А.О. Шамустакимова*

Генетика человека

- 80 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Ассоциации SAG-полиморфизма гена андрогенового рецептора с уровнем стероидных гормонов и антропометрическими показателями у мужчин: роль этнического фактора. *Л.В. Осадчук, Г.В. Васильев, А.В. Осадчук*
- 90 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Генетическая история коряков и эвенов Магаданской области по данным о полиморфизме Y-хромосомы. *Б.А. Малярчук, М.В. Деренко*

Генетика животных

- 98 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Универсальная панель STR-локусов для исследования полиморфизма вида *Canis lupus* и криминалистической идентификации собаки и волка. *А.Е. Гребенчук, И.С. Цыбовский*
- 108 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Сравнительные особенности геномного разнообразия кур *Gallus gallus domesticus* с декоративным фенотипом оперения «баки и борода». *Н.В. Деметьева, Ю.С. Щербаков, А.Е. Рябова, А.Б. Вахрамеев, А.В. Макарова, О.А. Николаева, А.П. Дысин, А.И. Азовцева, Н.Р. Рейнбах, О.В. Митрофанова*
- 117 **ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ**
Ассоциация трех однонуклеотидных полиморфизмов в гене *LPIN1* с показателями молочной продуктивности у коров ярославской породы. *А.В. Игошин, Т.М. Мишакова, Р.Б. Айтназаров, А.В. Ильина, Д.М. Ларкин, Н.С. Юдин*

Proceedings of the conference
“Resistance of plants and microorganisms
to adverse environmental factors”, Irkutsk

- 5 ORIGINAL ARTICLE
Reproductive relationships between taxa morphologically close to *Elymus caninus* (Poaceae: Triticeae). E.V. Shabanova, A.V. Agafonov, O.V. Dorogina
- 15 ORIGINAL ARTICLE
The regulatory effects of (p)ppGpp and indole on cAMP synthesis in *Escherichia coli* cells. N.M. Kashevarova, E.A. Khaova, A.G. Tkachenko
- 24 ORIGINAL ARTICLE
Effects of polyamines and indole on the expression of ribosome hibernation factors in *Escherichia coli* at the translational level. E.A. Khaova, A.G. Tkachenko
- 33 ORIGINAL ARTICLE
Expression of auxin transporter genes in flax (*Linum usitatissimum*) fibers during gravity response. N.N. Ibragimova, N.E. Mokshina
- 44 ORIGINAL ARTICLE
Ascorbate-glutathione cycle in wheat and rice seedlings under anoxia and subsequent reoxygenation. V.V. Yemelyanov, E.G. Prikaziuk, V.V. Lastochkin, O.M. Aresheva, T.V. Chirkova
- 55 ORIGINAL ARTICLE
Morphological and molecular analysis of rose cultivars from the Grandiflora and Kordesii garden groups. S.S. Yudanov, O.V. Dorogina, O.Yu. Vasilyeva

Plant genetics

- 63 ORIGINAL ARTICLE
Manifestation of agronomically valuable traits in the progeny of a sorghum mutant carrying the genetic construct for RNA silencing of the γ -kafirin gene. L.A. Elkonin, N.V. Borisenko, T.E. Pylaev, O.A. Kenzhegulov, S.Kh. Sarsenova, N.Yu. Selivanov, V.M. Panin
- 74 METHODS AND PROTOCOLS
Halo-RPD: searching for RNA-binding protein targets in plants. A.O. Shamustakimova

Human genetics

- 80 ORIGINAL ARTICLE
Associations of CAG repeat polymorphism in the androgen receptor gene with steroid hormone levels and anthropometrics among men: the role of the ethnic factor. L.V. Osadchuk, G.V. Vasiliev, A.V. Osadchuk
- 90 ORIGINAL ARTICLE
Genetic history of the Koryaks and Evens of the Magadan region based on Y chromosome polymorphism data. B.A. Malyarchuk, M.V. Derenko

Animal genetics

- 98 ORIGINAL ARTICLE
A universal panel of STR loci for the study of polymorphism of the species *Canis lupus* and forensic identification of dog and wolf. A.E. Hrebianchuk, I.S. Tsybovsky
- 108 ORIGINAL ARTICLE
Comparative peculiarities of genomic diversity in *Gallus gallus domesticus* chickens with decorative plumage: the muffs and beard phenotype. N.V. Dementieva, Y.S. Shcherbakov, A.E. Ryabova, A.B. Vakhrameev, A.V. Makarova, O.A. Nikolaeva, A.P. Dysin, A.I. Azovtseva, N.R. Reinbah, O.V. Mitrofanova
- 117 ORIGINAL ARTICLE
Association of three single nucleotide polymorphisms in the *LPIN1* gene with milk production traits in cows of the Yaroslavl breed. A.V. Igoshin, T.M. Mishakova, R.B. Aitnazarov, A.V. Ilina, D.M. Larkin, N.S. Yudin

Original Russian text <https://vavilovj-icg.ru/>

Reproductive relationships between taxa morphologically close to *Elymus caninus* (Poaceae: Triticeae)

E.V. Shabanova , A.V. Agafonov, O.V. Dorogina

Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

 ekobozeva87@mail.ru

Abstract. A hybridological study of biotypes of species close to *Elymus caninus*: *E. prokudinii*, *E. viridiglumis*, *E. goloskokovii*, as well as a number of morphologically deviant biotypes in Russia and Kazakhstan, was carried out. The objectives were to study the levels of reproductive relationships and the degree of integration of the species *E. goloskokovii*, *E. prokudinii*, and *E. viridiglumis* into the *E. caninus* complex. Our estimates of the seed fertility of natural parental biotypes were within 60–90 %. Among the combinations of crossing in F_1 , the highest seed setting was found in the hybrids formed by parental pairs from close habitats, regardless of the taxonomic rank of biotypes. The highest fertility values (55.6 and 46.1 %) were found in combinations involving *E. caninus*, *E. viridiglumis* and *E. goloskokovii*. It has been concluded that the biotypes of these species included in sexual hybridization form a single recombination gene pool, within which slight differences in reproductive compatibility are observed. The nature of the inheritance of the diagnostic features of lemmas “presence of trichomes” and “length of awns”, according to the digenic and monogenic type, respectively, is shown. The high seed fertility of the created hybrids and the presence of intermediate forms in the F_2 generation according to distinctive features indicate the possibility of interspecific introgression when species grow together in natural populations. Thus, the assessment of the inheritance of diagnostic characters makes it possible to classify *E. goloskokovii*, *E. prokudinii*, and *E. viridiglumis* as intraspecific taxa of *E. caninus* s. l. Data were obtained on the morphological and reproductive properties of interspecific hybrids with the participation of the species *E. mutabilis* as a possible donor in the speciation of taxa close to *E. caninus*. In cross combinations of *E. caninus* × *E. mutabilis* and *E. mutabilis* × *E. caninus*, lower values of seed fertility of hybrids in the F_1 and F_2 generations were noted compared to hybrids between the species *E. caninus*, *E. goloskokovii*, *E. prokudinii* and *E. viridiglumis*. Nevertheless, on the basis of chorological and morphological criteria, we concluded that *E. caninus* and *E. mutabilis* are independent species.

Key words: speciation; hybridization; inheritance; taxonomy; *Elymus*; Poaceae.

For citation: Shabanova E.V., Agafonov A.V., Dorogina O.V. Reproductive relationships between taxa morphologically close to *Elymus caninus* (Poaceae: Triticeae). *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(1):5-14. DOI 10.18699/vjgb-24-02

Репродуктивные взаимоотношения между представителями таксонов, морфологически близких к *Elymus caninus* (Poaceae: Triticeae)

Е.В. Шабанова , А.В. Агафонов, О.В. Дорогина

Центральный сибирский ботанический сад Сибирского отделения Российской академии наук, Новосибирск, Россия

 ekobozeva87@mail.ru

Аннотация. Проведено гибридологическое изучение биотипов видов, близких к *Elymus caninus*: *E. prokudinii*, *E. viridiglumis*, *E. goloskokovii*, а также ряда морфологически отклоняющихся биотипов с территории России и Казахстана. Была поставлена задача изучить уровни репродуктивных взаимоотношений и степени интегрированности видов *E. goloskokovii*, *E. prokudinii* и *E. viridiglumis* в комплекс *E. caninus*. Значения семенной фертильности природных родительских биотипов оценены нами в пределах 60–90 %. Среди комбинаций скрещивания в F_1 наивысшая завязываемость семян выявлена у гибридов, образованных родительскими парами из близких местообитаний, независимо от таксономической принадлежности биотипов. Наивысшие величины фертильности (55.6 и 46.1 %) обнаружены в комбинациях с участием видов *E. caninus*, *E. viridiglumis* и *E. goloskokovii*. Сделан вывод, что биотипы названных видов, включенные в половую гибридизацию, образуют единый рекомбинационный генпул, внутри которого наблюдаются незначительные различия по репродуктивной совместимости. Показан характер наследования диагностических признаков нижних цветковых чешуй «наличие трихом» и «длина остей» по дигенному и моногенному типу соответственно. Высокая семенная фертильность созданных гибридов и наличие в поколении F_2 промежуточных форм по различным признакам свидетельствуют о возможности межвидовой интрогрессии при совместном произрас-

тании видов в природных популяциях. Таким образом, оценка наследования диагностических признаков позволяет классифицировать *E. goloskokovii*, *E. prokudinii* и *E. viridiglumis* как внутривидовые таксоны *E. caninus* s. l. Получены данные о морфологических и репродуктивных свойствах межвидовых гибридов с участием вида *E. mutabilis* как возможного донора при видообразовании таксонов, близких к *E. caninus*. В комбинациях скрещивания *E. caninus* × *E. mutabilis* и *E. mutabilis* × *E. caninus* отмечены более низкие значения семенной фертильности гибридов в поколениях F₁ и F₂ по сравнению с гибридами между видами *E. caninus*, *E. goloskokovii*, *E. prokudinii* и *E. viridiglumis*. Тем не менее на основе хронологического и морфологического критериев нами сделан вывод о видовой самостоятельности *E. caninus* и *E. mutabilis*.

Ключевые слова: видообразование; гибридизация; признак; наследование; таксономия; *Elymus*; Poaceae.

Introduction

Wild cereals of the tribe Triticeae Dumort. (fam. Poaceae Barn.) are of great interest to researchers as possible donors of valuable traits for the main grain crops – wheat, barley and rye. General ideas about the potential possibilities of using wild relatives of wheat to enrich breeding material with new hereditary qualities were first outlined by N.I. Vavilov (1931). Subsequently, the concept of primary, secondary and tertiary gene pools was introduced (Harlan, De Wet, 1971). A similar gene pool system proposed for barley and rye (von Bothmer et al., 1992) includes, in general, some species of perennial grasses.

The genus *Elymus* L. is the largest genus of the tribe, belonging to the tertiary gene pool, uniting allopolyploid species of perennial grasses with different genomic constitutions. Facultative self-pollination, which promotes the elimination of spontaneous mutations and the consolidation of the consequences of introgressive hybridization, accelerates the processes of morphogenesis and at the same time complicates the systematization of natural populations. In the ongoing micro-evolutionary differentiation of the genus, the most relevant to study are the phylogenetic relationships between taxa.

Elymus caninus (L.) L. is a species with a StStHH genome (Dewey, 1968) and a vast range covering all of Europe from Iceland and the Mediterranean Sea to the Ural Mountains, almost the entire Palearctic part of Siberia, as well as some areas of Central Asia (Tsvelev, 1976; Hultén, Fries, 1986). In Northern Europe, *E. caninus* is distributed throughout Sweden and Denmark, and somewhat less frequently in Norway and Finland. In Siberia, it is found in almost all areas west of Lake Baikal (Peshkova, 1990). Taking into account the wide distribution of this species and its high adaptability to environmental factors, one can initially assume a noticeable variation in morphological characters within *E. caninus*. However, *E. caninus* exhibits low morphological variability compared to other *Elymus* species. The main reason, in our opinion, is that a number of natural morphotypes with deviating characters have been described as independent species. At the same time, on the one hand, in most cases no evidence was provided for the phylogenetic isolation of new species, and on the other hand, these species were not considered as a component of *E. caninus*, since they went beyond the limits of its narrow variability. In Russia and Kazakhstan, these species include *E. viridiglumis* (Nevski) Czer., *E. prokudinii* (Seregin) Tzvelev (Tsvelev, Probatova, 2019) and *E. goloskokovii* Kotuch. (Kotukhov, 2004).

The species *E. goloskokovii* was described from Western Altai (Ivanovsky Range), indicating its wide distribution within the southwestern part of these mountains (Kotukhov,

2004). The protologue notes that *E. goloskokovii* is a stable fertile hybridogenic species, probably derived from the hybridization of *E. fibrosus* (Schrenk) Tzvelev and *E. trachycaulis* (Link) Gould et Shinnars, with the possible participation of *E. mutabilis* (Drob.) Tzvelev. At the same time, the species *E. goloskokovii* differs from the widespread *E. caninus* mainly in the character of short (up to 4 mm) awns of lemmas.

Elymus viridiglumis was described from the Southern Urals in 1934 based on the collections of S.A. Nevsky as *Roegneria viridiglumis* Nevski. The species is distributed in the Urals and Western Siberia; it differs morphologically from *E. caninus* in having hairy or scabrous lemmas. Small populations have also been found in Eastern Kazakhstan.

Elymus prokudinii is an endemic of the subalpine meadows of the forest belt of the Central and Eastern Caucasus, described in 1965 as *Roegneria prokudinii* Seregin (Seregin, 1965) based on the collections of R.A. Elenevsky. The species is morphologically similar to *E. viridiglumis*, and differs from it only in its narrow endemic geographic range.

To date, much information has been accumulated indicating that the single recombination gene pool of *E. caninus*, as a species, is formed not only by typical individuals, but also by a large number of morphologically deviant biotypes (MDBs) that do not correspond to the diagnosis of the species (Gerus, Agafonov, 2006; Agafonov, 2011). In particular, we have obtained evidence that introgressive relationships between *E. caninus* and *E. mutabilis* lead to a diversity of transitional interspecific forms (Agafonov, 2013). Since that time, a number of questions remain and new problems have arisen from the perspective of reproductive biology and taxonomy of this vast complex.

One of the most important criteria for the relatedness of living organisms is the ability to produce viable offspring during crossing, which is due to the balanced recombination of genetic material during generations. The study of intra- and interspecific crossbreeding of biotypes makes it possible to model the processes of hybridization and introgression occurring within the genus *Elymus*. The results obtained from crossing biotypes from close or distant populations make it possible to clarify issues of intraspecific organization, outline the genetic pool and predict the possible course of further speciation pathways. The levels of crossbreeding of biotypes Cs (sexual compatibility) and the fertility of the resulting hybrids are under strict genotypic control and, accordingly, reflect the phylogenetic relationships of the original taxa (Agafonov et al., 2001).

In order to obtain additional data, work was carried out to create and study F₁–F₂ hybrids between selected biotypes of different taxa of species rank, morphologically close to

E. caninus s. l. The objectives were to study the levels of reproductive compatibility of biotypes and the degree of integration of the previously described species *E. goloskokovii*, *E. prokudinii* and *E. viridiglumis* into the *E. caninus* complex, as well as to supplement data on the morphological and reproductive properties of interspecific hybrids *E. caninus* × *E. mutabilis*.

Materials and methods

According to G.A. Peshkova (1990), the main diagnostic characteristics of *E. caninus* include: (1) leaf blades on top with scattered long hairs; (2) glabrate lemmas, rarely with single spines in the upper part; (3) lemmas with straight awns, equal to the lemmas or longer; (4) hairy rachillas.

In addition to the typical morphotype of *E. caninus* GAT-9210 and some morphologically deviating biotypes (MDB), biotypes of the above-mentioned species were used in hybridization in eight cross-combinations.

Characteristics by which typical individuals of *E. caninus* and *E. mutabilis* differ are as follows: long (up to 25 mm) – short (up to 6 mm) lemma’s awns; glabrous – scabrous (hairy) lemmas; the ratio of glume’s length and the adjacent lemma’s length ($k = L_{Gl}/L_{Lem}$) is approximately 0.5–0.6 in *E. caninus* and 0.7–0.8 in *E. mutabilis*. The last trait must be accompanied

by the presence of membranes at the edges of glumes, which become thinner with increasing k value (the *mutabilis* type), and conversely, membranes become wider as the k value decreases (the *caninus* type). This trait is not always clearly identified in most phenotypes due to the presence of a spectrum of intermediate phenotypes in natural populations. The locations of *E. mutabilis* accessions and species closely related to *E. caninus* are given in Table 1.

Fragments of flowering spikes of some biotypes of species close to *E. caninus* are shown in Figure 1. The chasmogamous type of flowering characteristic of all taxa does not prevent the predominant self-pollination of plants, which is supported by the simultaneous maturation of male and female gametophytes, as well as the absence of genetic systems of self-incompatibility (open (burst) anthers are visible in Fig. 1).

In addition to the diagnostic character “length of the awns of lemmas”, the biotypes included in the hybridization were found to have variability in a number of secondary characters inherent in specific taxa and geographic races: the relative length of the glumes, pubescence of leaf blades (LB), color of anthers, spike density, plant habit and height.

Since to study reproductive compatibility we use seeds of wild plants collected from different points of the species’ ranges, where their morphometric characters largely depend

Table 1. Localization of *E. mutabilis* accessions and species close to *E. caninus*, and their morphological features

Taxon, Accession	Locality and Collectors	Morphological features
<i>E. caninus</i> GAT-9210	Russia, the Altai Republic, northern coast of the Teletskoye lake, vicinity of the settl. Artybash, alt. 460 m N 51°47.674' E 87°16.446' (V. Makashov)	Lemmas glabrous; Leaf blades hairy on upper surfaces; Rachilla segments hairy
<i>E. caninus</i> MDB AKL-0703	Russia, the Altai region, vicinity of the settl. Blagoveshchenka, alt. 98 m N 52°55.19' E 79°46.22' (N. Lashchinsky)	Lemmas hairy; Leaf blades hairless
<i>E. caninus</i> MDB SON-9904	Russia, the Republic of Khakassia, Zapadnyy Sayan range, 252 km of the Abaza–Ak-Dovurak highway, alt. 713 m N 52°10.772' E 89°51.907' (A. Agafonov)	Lemmas short and hairy; Lemma’s awns up to 12 mm; Leaf blades hairy on upper surfaces
<i>E. caninus</i> MDB OSE-1427	Russia, the Republic of North Ossetia – Alania, vicinity of the settl. Stur-Digora, alt. 1996 m N 42°52.898' E 43°35.959' (S. Asbaganov)	Rachilla segments scabrous due to short spinules
<i>E. caninus</i> MDB BAI-0401	Russia, the Republic of Buryatia, vicinity of the settl. Goryachinsk, coast of Lake Baikal, alt. 457 m N 52°58.588' E 108°16.335' (D. Gerus, A. Agafonov)	Leaf blades scabrous
<i>E. viridiglumis</i> EK-1418	The Republic of Kazakhstan, vicinity of the settl. Berezovka, alt. 1202 m N 50°07.623' E 83°49.210' (D. Gerus)	Lemmas scabrous; Leaf blades hairy on upper surfaces
<i>E. viridiglumis</i> BEL-1404	Russia, the Altai region, vicinity of the settl. Belokurikha, alt. 287 m N 51°58.847' E 84°57.697' (A. Agafonov, M. Agafonova)	Lemmas short and hairy; Leaf blades hairy on upper surfaces
<i>E. goloskokovii</i> TUV-9936	Russia, the Republic of Tuva, Todzhinsky district, floodplain of the Biy-Khem River, Tos-Buluk shoal (D. Shaulo)	Lemma’s awns up to 3 mm; Leaf blades hairy on upper surfaces
<i>E. goloskokovii</i> EK-1513	The Republic of Kazakhstan, vicinity of the settl. Poperechnoe, alt. 1202 m N 50°21.128' E 83°53.527' (D. Gerus)	Lemma’s awns up to 4 mm; Leaf blades hairless
<i>E. prokudinii</i> TEB-1806	Russia, the Karachay-Cherkess Republic, the Teberda Nature Reserve, alt. 2020 m N 43°26.508' E 41°42.693' (S. Asbaganov, A. Agafonov)	Lemmas short and hairy; Leaf blades hairy on upper surfaces
<i>E. mutabilis</i> BAI-0402	Russia, the Republic of Buryatia, vicinity of the settl. Goryachinsk, shore of Lake Baikal, alt. 457 m N 52°58.588' E 108°16.335' (D. Gerus, A. Agafonov)	Lemmas scabrous; Lemma’s awns up to 4 mm
<i>E. mutabilis</i> ACH-8932	Russia, the Altai Republic, Cheke-Taman pass, alt. 1216 m N 50°38.367' E 86°18.166' (A. Agafonov)	Lemmas scabrous; Lemma’s awns up to 2 mm

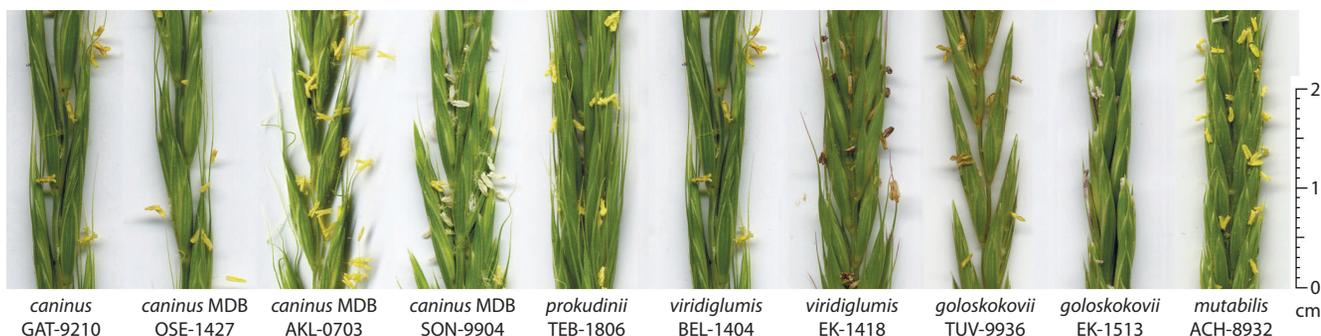


Fig. 1. Fragments of flowering spikes of taxa and biotypes close to *E. caninus*, taken for hybridization.

on environmental factors, it is necessary to exclude modification of taxonomic characters. To do this, plants were grown under equalized conditions at the experimental site of the Central Siberian Botanical Garden SB RAS, and only after that their taxonomic affiliation was determined. When selecting parental individuals, forms with different characteristics were selected – plants with glabrous lemmas were crossed with plants that had trichomes on lemmas, plants with short awns of lemmas were crossed with plants with long awns, etc.

The procedures for creating sexual hybrids were carried out using an express method (Lu et al., 1990), which involves preliminary preparation of the ear and manual pollination of each pistil. One of the advantages of this technique is the stimulation of the natural opening of flowers and the simultaneous removal of anthers that have not yet burst. Pollination of each of the several opened emasculated flowers is carried out by the newly burst anther of the father plant, which minimizes the risk of its own or foreign pollen. The use of this technique, with sufficient development, gives more reliable results, since it does not require preliminary emasculation of delicate immature flowers and leads to an increase in the efficiency of hybridization.

Hybridity of F_1 plants was confirmed by the presence of characteristics of the paternal plant. The assessment of seed fertility (SF) of plants in generations F_1 – F_2 and levels of sexual

compatibility of biotypes (Cs) was carried out according to the principles we developed (Agafonov, 1994; Agafonov, Salomon, 2002). The correspondence of the type of inheritance of morphological characters (presence of trichomes on lemmas and length of lemma's awns) to Mendel's laws (monogenic and digenic) in F_2 plants was checked using the Pearson criterion (χ^2) (Pearson, 1900).

The segregation of the awn length trait in F_2 plants was analyzed based on the maximum value of awn length among the spikes of each individual.

Results

Crosses were carried out between species close to *E. caninus* in six combinations: *E. caninus* × *E. prokudinii* (2 combinations), *E. caninus* × *E. viridiglumis*, *E. goloskokovii* × *E. caninus*, *E. caninus* × *E. goloskokovii*, *E. goloskokovii* × *E. viridiglumis*; there were also two cross combinations between the biotypes *E. caninus* and *E. mutabilis*. From 1 to 3 hybrid grains were obtained in each combination. The results are presented in Table 2.

The inheritance of morphological characters was assessed in small samples of the F_2 generation. Although the sample sizes were not large, they made it possible to assess the degree of discreteness of the trait in phenotypic classes and the nature of their inheritance (Tables 3, 4).

Table 2. Maximum values of seed fertility in F_1 and F_2 hybrids (%) and levels of sexual compatibility (Cs) of biotypes among taxa close to *E. caninus* and in cross combinations of *E. caninus* × *E. mutabilis*

No.	Cross combination	N_{F_1}	F_1 (max)	N_{F_2}	F_2 (max)	Level of Cs
1	<i>caninus</i> MDB AKL-0703 × <i>prokudinii</i> TEB-1806	2	24.4	18	88.8	α_1
2	<i>goloskokovii</i> EK-1513 × <i>viridiglumis</i> EK-1418	1	46.1	15	77.1	α_1
3	<i>goloskokovii</i> EK-1513 × <i>caninus</i> MOB SON-9904	3	29.6	125	91.3	α_1
4	<i>caninus</i> MDB SON-9904 × <i>goloskokovii</i> TUV-9936	2	55.6	82	63.6	α_1
5	<i>caninus</i> MDB OSE-1427 × <i>prokudinii</i> TEB-1806	1	39.7	16	79.4	α_1
6	<i>caninus</i> MDB OSE-1427 × <i>viridiglumis</i> BEL-1404	2	10/7*	6	79.0	α_2
7	<i>caninus</i> GAT-9210 × <i>mutabilis</i> ACH-8932	1	6.6	13	54.6	α_2
8	<i>mutabilis</i> BAI-0402 × <i>caninus</i> BAI-0401	1	3.2	14	28.6	α_2

Note. N_{F_1} and N_{F_2} are the number of analyzed plants in F_1 and F_2 , respectively.

*The fraction shows the ratio of the number of completed seeds to the number of spikes from two F_1 plants.

Table 3. Results of trait segregation in F₂ in small samples

Crossing combination	Trait	Trait expression	Number of plants	Total plants in F ₂
<i>caninus</i> AKL-0703 × <i>prokudinii</i> TEB-1806	Trichomes on leaf blades	Scabrous	6	18
		Hairy	12	
<i>goloskokovii</i> EK-1513 × <i>viridiglumis</i> EK-1418	Trichomes on lemmas	Glabrous	4	15
		Scabrous	11	
	Lemma's awn length	1–3 mm	4	
		5–8 mm	8	
		18 mm	3	
<i>caninus</i> OSE-1427 × <i>prokudinii</i> TEB-1806	Trichomes on lemmas	Glabrous	5	16
		Short hairy	11	
	Trichomes on leaf blades	Scabrous	1	
		Hairy	15	
	Trichomes on rachillas	Short spiny	4	
		Hairy	12	
<i>caninus</i> OSE-1427 × <i>viridiglumis</i> BEL-1404	Trichomes on lemmas	Glabrous	4	6
		Scabrous	2	
	Trichomes on leaf blades	Scabrous	5	
		Short hairy	1	
	Trichomes on rachillas	Short spiny	3	
		Hairy	3	
<i>caninus</i> GAT-9210 × <i>mutabilis</i> ACH-8932	Trichomes on lemmas	Glabrous	2	13
		Scabrous	3	
		Short hairy	8	
	Trichomes on leaf blades	Scabrous	2	
		Hairy	11	
	Length of lemma's awns	1–4 mm	7	
		6–8 mm	1	
		10–12 mm	2	
		18 mm	1	
		19–22 mm	2	
		$k = L_{GI}/L_{Lem}$	0.5–0.6	
		0.6–0.7	8	
		0.7–0.8	3	
	0.8–0.9	1		
<i>mutabilis</i> BAI-0402 × <i>caninus</i> BAI-0401	Trichomes on lemmas	Glabrous	3	14
		Scabrous	7	
		Short spiny	4	
	Trichomes on leaf blades	Scabrous	6	
		Hairy	8	
	Length of lemma's awns	1–3 mm	5	
		4–6 mm	3	
		7–8 mm	4	
		18–20 mm	2	
		$k = L_{GI}/L_{Lem}$	0.5–0.6	
		0.7–0.8	3	
	0.8–0.9	4		

Table 4. Analysis of trait segregation in the F₂ generation in the two largest samples

Cross combination	Trait	Trait expression	p*	q*	χ ² , P
<i>goloskokovii</i> EK-1513 × <i>caninus</i> SON-9904	Trichomes on lemmas 1:15	Glabrous	8	7.8125	χ ² = 0.0048
		Scabrous, densely scabrous, and short and hairy	117	117.1875	(0.95 > P > 0.90)
Total 125 plants F ₂	Length of lemma's awns 1:2:1	1–4 mm	34	31.25	χ ² = 8.856
		6–9 mm	74	62.5	(0.01 > P > 0.025)
		10–12 mm	17	31.25	
<i>caninus</i> SON-9904 × <i>goloskokovii</i> TUV-9936	Trichomes on lemmas 1:15	Glabrous	7	5.125	χ ² = 0.732
		Scabrous, densely scabrous, and short and hairy	75	76.875	(0.90 > P > 0.10)
Total 82 plants F ₂	Length of lemma's awns 1:2:1	1–3 mm	17	20.5	χ ² = 3.15
		6–9 mm	49	41.0	(0.90 > P > 0.10)
		10–12 mm	16	20.5	

* p – actual number of individuals; * q – expected number of individuals.

Pearson's test χ^2 was applied and the level of significance P was assessed for F₂ samples in two cross combinations: *E. goloskokovii* × *E. caninus* and *E. caninus* × *E. goloskokovii* (see Table 4).

Discussion

The results of the study showed that the biotypes included in the hybridization, which are close in morphology to the base species *E. caninus* (see Table 2, No. 1–6), form a single recombination gene pool, within which minor differences in reproductive compatibility are observed. Noteworthy is the relatively small but clear decrease in SF values in F₁ hybrids compared to mother plants. This means that all MDBs taken for hybridization belong to the extensive gene pool of *E. caninus*, but have gone through a certain microevolutionary path in the direction of divergence.

We estimated the SF values of natural biotypes to be within 60–90 %. The maximum values of SF of F₁ plants among species close to *E. caninus* (see Table 2, No. 1–5) were in the range of 24.4–55.6 %, except for the crossing combination of *E. caninus* × *E. viridiglumis* (No. 6), where SF was very low; for this reason, the maximum value of SF could not be given. In this combination, out of seven spikes (the total number of spikes in two F₁ plants), only 10 seeds were fertile, of which only six were able to grow to the generative stage and produce offspring in F₂. In cross combinations between the species *E. caninus* and *E. mutabilis* (see Table 2, No. 7, 8), where one F₁ plant was obtained, the maximum SF value was in the range of 3.2–6.6 %.

The highest SF values in F₁ were found in hybrid plants formed by parental pairs from geographically close habitats (see Table 2): combination 4 from the border territory of the Republic of Khakassia and the Republic of Tuva (*E. caninus* SON-9904 × *E. goloskokovii* TUV-9936) – 55.6 %, and combination 2 from North-Eastern Kazakhstan (*E. goloskokovii* EK-1513 × *E. viridiglumis* EK-1418) – 46.1 %. Combination 5

from the North Caucasus (*E. caninus* OSE-1427 × *E. prokudinii* TEB-1806) was also characterized by a relatively high SF – 39.7 %. This fact also confirms our assumption about the joint microevolutionary path of different taxa within a specific territory (Agafonov, 2011). Moreover, two interspecific combinations (7 and 8), formed by *E. caninus* and *E. mutabilis*, showed significantly lower (6.6 and 3.2 %) SF values in the F₁ generation than other combinations (1–5). As for the hybridization variant *E. caninus* OSE-1427 × *E. viridiglumis* BEL-1404 (6), low SF values in the F₁ hybrid can be explained by the significant geographic isolation of the original parental biotypes.

Generally, the values of seed fertility in F₂ represent a range of variability within certain limits, determined, among other things, by the degree of phylogenetic proximity of the parental biotypes. In all crossing combinations, an increase in SF values in F₂ was observed (see Table 2), which is associated with the normalization of genetic recombination according to the RGP principle. The levels of sexual compatibility of biotypes in almost all crossing variants (except for two involving *E. mutabilis* and variant 6) correspond to free genetic recombination (α_1). The possibility of interbreeding and restoration of seed productivity over several generations indicates the presence of a homologous part of the genome in the parental forms. Such hybridization can be considered as natural seed propagation of two different forms of the same species.

Based on the presented results of hybridization, it was concluded that the nature of inheritance of traits during the segregation of hybrids in F₂ in self-pollinating taxa depends on the parental genotypes, which can differ in a different number of loci and alleles (one or more), forming the general gene pool of the taxon. In these crossing combinations, the nature of inheritance of the trait “presence of trichomes on lemmas” was determined to be digenic, while the inheritance of the trait “awn length” was determined to be monogenic with incomplete dominance.

The distribution of phenotypes according to the trait “presence of trichomes on lemmas” by class allowed us to assume a segregation of 1:15, which corresponds to digenic inheritance. The level of significance test (P) was low in some cases, which may be explained by the insufficient number of F₂ plants forming the sample. However, based on the segregation of phenotypes, one can roughly assume the type of inheritance of diagnostic traits.

Previously, we conducted expeditionary collections of species close to *E. caninus* in the Republics of Altai and Tuva, the Caucasus, and Eastern Kazakhstan. The samples were studied experimentally. Let us give a brief overview of the most significant results for this article. Since small populations of *E. viridiglumis* were found to occur in Eastern Kazakhstan, we hypothesized possible routes for the formation of the species in this territory (Agafonov, 2013). One cannot but agree with the comment that according to spike characteristics, some biotypes of *E. viridiglumis* are similar to *E. komarovii* (Tsvelev, Probatova, 2019); however, the genetic distance of these two species was experimentally shown (Agafonov et al., 2017). A taxon similar to *E. goloskokovii* in the character “short awns of lemmas” was described from Northern Europe as *E. caninus* var. *muticus* (Holmb.) Karlsson. We studied the living material of this taxon in an experiment (Gerus, Agafonov, 2006), and it was concluded that this species has a polyphyletic hybrid origin.

ISSR analysis using a wide range of samples of species morphologically similar to *E. caninus* from different localities within Russia showed that the species *E. viridiglumis*, *E. prokudinii* and *E. goloskokovii* represent groups of individuals that are also phylogenetically close to *E. caninus* (Shabanova (Kobozeva) et al., 2020). The assumption was confirmed that *E. viridiglumis* has a polyphyletic origin, as a result of microevolutionary processes in populations of *E. caninus* s. l., possibly with the participation of *E. mutabilis*. For the Caucasian endemic *E. prokudinii* and the Kazakh endemic *E. goloskokovii*, origin is also assumed to be a result of introgression or spontaneous mutagenesis, i. e. manifestations of natural intraspecific polymorphism of *E. caninus*. The remoteness of *E. fibrosus* from all taxa phylogenetically close to *E. caninus* cast doubt on the assumption of the origin of *E. goloskokovii* from the hybridization of *E. fibrosus* and *E. trachycaulis*, especially considering the introduced North American origin of the latter (Agafonov, Baum, 2000). Taking into account our early studies of species close to *E. caninus* (Agafonov, 2011), it was concluded that the taxa *E. viridiglumis*, *E. prokudinii* and *E. goloskokovii* are not phylogenetically separate and should be transferred to the intraspecific rank of *E. caninus* s. l.

Reproductive relationships between *Elymus caninus* and *E. mutabilis*

The typical morphotypes of *E. caninus* and *E. mutabilis* have a special character of reproductive relationships. Previously, we touched upon the topic of the influence of *E. mutabilis* on speciation in Eastern Kazakhstan (Gerus, Agafonov, 2006), including in comparison with data on intraspecific variability of *E. caninus* (Agafonov, 2011). Some *E. caninus* × *E. caninus* hybrids were found to have lower compatibility than *E. caninus* × *E. mutabilis* hybrids.

In this study, the parental biotypes of *E. caninus* and *E. mutabilis* had typical characteristics of these species; differences were identified only between biotypes of *E. caninus* in the pubescence of the upper surfaces of leaf blades: hairy surfaces were noted in the parental biotype GAT-9210, scabrous (hairless) ones were noted in the BAI-0401 biotype. Data on SF and the segregation of morphological traits' characters in F₂ hybrid samples in two cross combinations are given in Tables 5 and 6.

***Elymus caninus* GAT-9210 × *E. mutabilis* ACH-8932.** In the grown F₂ sample of 13 individuals (see Table 5), only one absolutely sterile plant (4) was noted; the highest SF value was 54.6 % (plant 8). This value was slightly lower than in all combinations involving taxa close to *E. caninus* (see Table 2). In addition, a wide range of SF values was noted, which is a characteristic feature for relatively distant hybrids. Based on a set of morphological characters, two individuals (2 and 8) were recombinant ones (see Table 5), and corresponded to the diagnosis of *E. viridiglumis* (hairy upper surfaces of leaf blades, scabrous lemmas and long awns up to 18–22 mm). This fact confirms our assumption about the polyphyletic origin of the abovementioned taxon. Moreover, only two individuals had scabrous upper surfaces of leaf blades (plants 7 and 13), the rest had hairy surfaces to a greater or lesser extent (11 individuals). This fact may indicate a small number of alleles by which the parental individuals differed, since during segregation in F₂ for one discriminating allele, there is a greater probability of obtaining 2 individuals with recessive homozygotes per sample of 13 plants (1 out of 4) than with two alleles (1 out of 16).

Based on the diagnostic trait of lemma's awn length, we identified morphotypes with six values from 3 mm (morphotype *E. mutabilis*) to 22 mm (morphotype *E. caninus*). The large number of phenotypes for this trait can probably be explained by the fact that each allele makes an additive contribution to the formation of the awn length trait. And the greater the difference in the length of the awns in the parental forms, the greater the range of variability in the offspring. In addition, the value of this quantitative characteristic is quite difficult to fix.

***Elymus mutabilis* BAI-0402 × *E. caninus* BAI-0401.** The F₂ sample in this combination is represented by 14 individuals with an average SF value of 11.0 % (see Table 6), which is the smallest value among all studied samples. At the same time, the maximum value of SF in F₂, 28.6 %, was also the smallest compared to other samples (see Table 2). In the sample, six individuals with completely naked leaf blades were noted (see Tables 3, 6). This may mean that the parental biotype of *E. mutabilis* BAI-0402 was heterozygous for the allele(s) controlling the trait. At the same time, three individuals with glabrous lemmas (morphotype *E. caninus*) differed from each other in the length of lemma's awns (see Table 6). That is, potentially new taxa have emerged if traditional classification methods are followed. Despite the fact that individual 13 was completely sterile, the other two (3 and 4) had, although reduced, quite sufficient seed fertility (9.4 and 7.0 %) for their own reproduction and consolidation in the next generations.

In a small part of the overlapping areas of *E. caninus* and *E. mutabilis* within the mountainous regions of the Republic

Table 5. Main morphological characteristics and seed fertility (SF) in the F₂ sample of the *E. caninus* GAT-9210 × *E. mutabilis* ACH-8932 hybrid

No.	Upper surfaces of leaf blades	$k = L_{Gl}/L_{Lem}$	Trichomes on lemmas	Length of lemma's awns, mm	SF, %
1	Hairy	0.7–0.8	Densely scabrous	4	3.4
2	Hairy	0.6–0.7	Scabrous	22	37.5
3	Hairy	0.6–0.7	Scabrous	4	5.5
4	Hairy	0.6–0.7	Densely scabrous	6	0
5	Hairy	0.7–0.8	Densely scabrous	4	39.7
6	Hairy	0.8–0.9	Densely scabrous	12	4.8
7	Scabrous	0.6–0.7	Glabrous	3	10.7
8	Hairy	0.6–0.7	Densely scabrous	22	54.6
9	Hairy	0.5–0.6	Densely scabrous	3	3.6
10	Hairy	0.6–0.7	Densely scabrous	12	12/5*
11	Hairy	0.7–0.8	Densely scabrous	4	14.2
12	Hairy	0.6–0.7	Scabrous	18	3.2
13	Scabrous	0.6–0.7	Glabrous	3	33.8
SF average value %					16.2

* The fraction shows the ratio of the number of completed seeds to the number of spikes from two F₁ plants.

Table 6. Main morphological characteristics and seed fertility (SF) in the F₂ sample of the *E. mutabilis* BAI-0402 × *E. caninus* BAI-0401 hybrid

No.	Upper surfaces of leaf blades	$k = L_{Gl}/L_{Lem}$	Trichomes on lemmas	Length of lemma's awns, mm	SF, %
1	Scabrous	0.5–0.6	Scabrous	3	12.2
2	Scabrous	0.5–0.6	Scabrous	8	24.5
3	Hairy	0.5–0.6	Glabrous	3	9.4
4	Hairy	0.7–0.8	Glabrous	20	7.0
5	Hairy	0.8–0.9	Densely scabrous	6	5.2
6	Hairy	0.5–0.6	Scabrous	8	0
7	Scabrous	0.5–0.6	Scabrous	6	7.8
8	Hairy	0.5–0.6	Scabrous	8	13.4
9	Scabrous	0.7–0.8	Densely scabrous	6	2.0
10	Hairy	0.8–0.9	Scabrous	3	19.7
11	Scabrous	0.8–0.9	Scabrous	3	28.6
12	Hairy	0.7–0.8	Densely scabrous	3	3.6
13	Scabrous	0.8–0.9	Glabrous	8	0
14	Hairy	0.5–0.6	Densely scabrous	20	21.4
SF average value %					11.0

of Khakassia (Krasnoyarsk Territory), we found biotypes with extreme and all intermediate values of the trait “lemma’s awn length” (Fig. 2). This phenomenon can only be explained by acts of mutual introgression between the two species, as well as multiple allelism of the genes that control this trait. The

existence of the introgression mechanism is confirmed by the SF values in interspecific hybrids (see Tables 5, 6).

The hybrids we created in *E. caninus* × *E. mutabilis* combinations had reduced seed fertility at the level of α_2 . This level of SF certainly reduces the competitive ability in natural con-



Fig. 2. Fragments of spikes in natural populations from Khakassia. Variation in the lemma's awns length in mixed populations of *E. mutabilis* and *E. caninus*.

1 – typical *E. mutabilis*; 7 – typical *E. caninus*; 2–6 – intermediate morphotypes.

ditions, but the probability of the formation of descendants and their consolidation in populations is quite high. In general, the increase in SF in *E. caninus* × *E. mutabilis* hybrids to a normal level already in the F₂ generation confirms the possibility of a fairly easy exchange of genetic material between species. This means that some spontaneous hybrids in natural conditions have a chance to survive in subsequent generations, while increasing the overall population biodiversity, as has been shown previously (Sun et al., 1999).

Conclusion

Thus, based on the indicators of interbreeding and character segregation among taxa close to the widespread species *E. caninus*, an integral assessment of the relationships between biotypes was obtained. From the results of the study, the feasibility of a taxonomic revision logically follows. In our opinion, the taxa currently recognized as independent species *E. prokudinii*, *E. viridiglumis* and *E. goloskokovii* due to their polyphyletic origin must be transferred to the intraspecific rank within the polymorphic species *Elymus caninus* s. l. The main confirmation of the phylogenetic unity of these taxa is the high values of SF already in the first generation of hybrids and free recombination of diagnostic characters (reproductive compatibility) at the α_1 level.

Based on the results of chorological analysis and hybridization of selected biotypes of *Elymus mutabilis* and *E. caninus*, it was concluded that *E. mutabilis* is an independent species with the widest range in the Northern Hemisphere and with high intraspecific variability in many characters.

References

- Agafonov A.V. The principle of recombination gene pools (RGP) and introgression gene pools (IGP) in the biosystematic treatment of *Elymus* species. In: Proc. of the 2nd Int. Triticeae Symp. Logan, Utah, USA, 20–24 June, 1994. Logan, 1994;254-260
- Agafonov A.V. The general structure of the recombination gene pool of *Elymus caninus* (Triticeae: Poaceae) according to the data on interbreeding and the assessment of the inheritance of some morphological characters used in taxonomy. *Rastitelnyj Mir Aziatskoj Rossii = Flora and Vegetation of Asian Russia*. 2011;2(8):61-70 (in Russian)
- Agafonov A.V. The single recombination gene pool of *Elymus caninus* and *E. mutabilis* (Triticeae: Poaceae) according to interbreeding and natural variability: taxonomic implications. In: Proceedings of the All-Russia Conference “The Flora of North Asia: Issues of Studying and Preserving Biodiversity”. Novosibirsk, October 1–3, 2013. Novosibirsk: Central Siberian Botanical Garden, Siberian Branch of the RAS, 2013;3-5 (in Russian)
- Agafonov A.V., Baum B.R. Individual variability and reproductive properties of sexual hybrids within the *Elymus trachycaulus* (Poaceae: Triticeae) complex and related taxa. 1. Polymorphism of endosperm storage proteins in North American and Eurasian biotypes. *Turczaninowia*. 2000;3(1):63-75 (in Russian)
- Agafonov A.V., Salomon B. Genepools among SH genome *Elymus* species in boreal Eurasia. In: Triticeae IV. Consejo de Agricultura y Pesca. Spain, Sevilla, 2002;37-41
- Agafonov A.V., Baum B.R., Bailey L.G., Agafonova O.V. Differentiation in the *Elymus dahuricus* complex (Poaceae): evidence from grain proteins, DNA, and crossability. *Hereditas*. 2001;135(2-3): 277-289. DOI 10.1111/j.1601-5223.2001.00277.x
- Agafonov A.V., Asbaganov S.V., Emtseva M.V., Kobozeva E.V. Reticulate microevolution in mixed populations *Elymus uralensis*, *E. viridiglumis*, *E. mutabilis*, and *E. caninus* (Poaceae) in Southern Urals according to ISSR-analysis. In: Proceedings of the 6th Int. Sci. Conf. “Issues of Studying the Vegetation Cover of Siberia”. Tomsk, October 24–26, 2017. Tomsk: Publ. House of the Tomsk State University, 2017;143-145. DOI 10.17223/9785946216371/46 (in Russian)
- Dewey D.R. Synthetic *Agropyron–Elymus* hybrids: *Elymus canadensis* × *Agropyron caninum*, *A. trachycaulum* and *A. striatum*. *Am. J. Bot.* 1968;55(10):1133-1139. DOI 10.2307/2440735
- Gerus D.E., Agafonov A.V. Biosystematic analysis of the origin of some taxa and morphologically deviating forms closely related to *Elymus caninus* and *E. mutabilis*. *Siberian Botanical Bulletin: electronic journal*. 2006;1(1):65-74 (in Russian)
- Harlan J.R., de Wet J.M.J. Toward a rational classification of cultivator plants. *Taxon*. 1971;20(4):509-517. DOI 10.2307/1218252
- Hultén E., Fries M. Atlas of North European Vascular Plant, North of the Tropic of Cancer. Lubrecht & Cramer Ltd., 1986
- Kotukhov Yu.A. New species of cereals (Poaceae) from Western Altai. *Turczaninowia*. 2004;7(4):8-10 (in Russian)
- Lu B.-R., Salomon B., von Bothmer R. Cytogenetic studies of progeny from the intergeneric crosses *Elymus* × *Hordeum* and *Elymus* × *Secale*. *Genome*. 1990;33(3):425-432. DOI 10.1139/g90-064
- Nevsky S.A. Genus 196. *Roegneria* C. Koch. In: Flora of the USSR. Vol. 2. Cereals. Leningrad: USSR Academy of Sciences, 1934;599-627 (in Russian)

- Nevsky S.A. List of cereals from the tribes Lolieae, Nardeae, Leptureae, and Hordeae of the USSR flora. In: Flora and Systematics of Higher Plants. Iss. 2. Moscow–Leningrad: Publ. House of the USSR Academy of Sciences, 1936;33-90 (in Russian)
- Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* 1900;50(302):157-175. DOI 10.1080/14786440009463897
- Peshkova G.A. *Elymus* L. – Couch Grass. In: Flora of Siberia. Vol. 2. Novosibirsk: Nauka Publ., 1990;17-32 (in Russian)
- Seredin R.M. A new species of the genus *Roegneria* C. Koch from the North Caucasus. *Novosti Systematiki Vysshykh Rasteniy = Novitates Systematicae Plantarum Vascularium.* 1965;2:55-57 (in Russian)
- Shabanova (Kobozeva) E., Emtseva M., Agafonov A., Lu B.-R. Assessment of the variability of *Elymus caninus* (Poaceae) and closely related taxa from Russia and Kazakhstan using ISSR markers. *BIO Web Conf.* 2020;24:00076. DOI 10.1051/bioconf/20202400076
- Sun G.-L., Diaz O., Salomon B., von Bothmer R. Genetic diversity in *Elymus caninus* as revealed by isozyme RAPD and microsatellite markers. *Genome.* 1999;42(3):420-431
- Tsvelev N.N. Cereals of the USSR. Leningrad: Nauka Publ., 1976 (in Russian)
- Tsvelev N.N., Probatova N.S. Cereals of Russia. Moscow, 2019 (in Russian)
- Vavilov N.I. Linnean species as a system. *Trudy po Prikladnoy Botanike, Genetike i Seleksii = Proceedings on Applied Botany, Genetics, and Breeding.* 1931;26(3):109-134 (in Russian)
- von Bothmer R., Seberg O., Jacobsen N. Genetic resources in the Triticeae. *Hereditas.* 1992;116(1):141-150. DOI 10.1111/j.1601-5223.1992.tb00814.x

ORCID

E.V. Shabanova (Kobozeva) orcid.org/0000-0002-6566-9325
A.V. Agafonov orcid.org/0000-0002-1403-5867
O.V. Dorogina orcid.org/0000-0001-5729-3594

Acknowledgements. The study was carried out within the framework of State Assignment of CSBG SB RAS No. AAAA-A21-121011290025-2 and was financially supported by the Ministry of Science and Higher Education of the Russian Federation under Agreement No. 075-15-2021-1056 of September 28, 2021 between BIN RAS and the Ministry of Science and Higher Education of the Russian Federation, also under Agreement No. ЕП/29-10-21-4 between BIN RAS and CSBG SB RAS. The materials of the bio-resource scientific collections of the CSBG SB RAS USU No. 440534 and USU No. 440537 (Herbaria NS, NSK) were used. The authors are grateful to engineer M.V. Emtseva for assistance in conducting hybridological studies.

Conflict of interest. The authors declare no conflict of interest.

Received June 8, 2023. Revised August 30, 2023. Accepted October 6, 2023.

Original Russian text <https://vavilovj-icg.ru/>

The regulatory effects of (p)ppGpp and indole on cAMP synthesis in *Escherichia coli* cells

N.M. Kashevarova , E.A. Khaova, A.G. Tkachenko

Institute of Ecology and Genetics of Microorganisms of the Ural Branch of the Russian Academy of Sciences, Perm, Russia
 nkashev@mail.ru

Abstract. Bacterial stress adaptive response is formed due to changes in the cell gene expression profile in response to alterations in environmental conditions through the functioning of regulatory networks. The mutual influence of network signaling molecules represented by cells' natural metabolites, including indole and second messengers (p)ppGpp and cAMP, is hitherto not well understood, being the aim of this study. *E. coli* parent strain BW25141 ((p)ppGpp⁺) and deletion knockout BW25141Δ*relA*Δ*spoT* which is unable to synthesize (p)ppGpp ((p)ppGpp⁰) were cultivated in M9 medium supplemented with different glucose concentrations (5.6 and 22.2 mM) in the presence of tryptophan as a substrate for indole synthesis and in its absence. The glucose content was determined with the glucose oxidase method; the indole content, by means of HPLC; and the cAMP concentration, by ELISA. The onset of an increase in initially low intracellular cAMP content coincided with the depletion of glucose in the medium. Maximum cAMP accumulation in the cells was proportional to the concentration of initially added glucose. At the same time, the (p)ppGpp⁰ mutant showed a decrease in maximum cAMP levels compared to the (p)ppGpp⁺ parent, which was the most pronounced in the medium with 22.2 mM glucose. So, (p)ppGpp was able to positively regulate cAMP formation. The promoter of the tryptophanase operon responsible for indole biosynthesis is known to be under the positive control of catabolic repression. Therefore, in the cells of the (p)ppGpp⁺ strain grown in the tryptophan-free medium that were characterized by a low rate of spontaneous indole formation, its synthesis significantly increased in response to the rising cAMP level just after glucose depletion. However, this was not observed in the (p)ppGpp⁰ mutant cells with reduced cAMP accumulation. When tryptophan was added to the medium, both of these strains demonstrated high indole production, which was accompanied by a decrease in cAMP accumulation compared to the tryptophan-free control. Thus, under glucose depletion, (p)ppGpp can positively regulate the accumulation of both cAMP and indole, while the latter, in its turn, has a negative effect on cAMP formation.

Key words: *Escherichia coli*; signaling molecules; cAMP; (p)ppGpp; indole; glucose; tryptophan.

For citation: Kashevarova N.M., Khaova E.A., Tkachenko A.G. The regulatory effects of (p)ppGpp and indole on cAMP synthesis in *Escherichia coli* cells. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024; 28(1):15-23. DOI 10.18699/VJGB-24-03

Регуляторные эффекты (p)ppGpp и индола на синтез цАМФ в клетках *Escherichia coli*

Н.М. Кашеварова , Е.А. Хаова, А.Г. Ткаченко

Институт экологии и генетики микроорганизмов Уральского отделения Российской академии наук – филиал Пермского федерального исследовательского центра Уральского отделения Российской академии наук, Пермь, Россия
 nkashev@mail.ru

Аннотация. Адаптивный ответ бактерий на стрессорное воздействие условий среды формируется за счет изменения генно-экспрессионного профиля клетки посредством функционирования регуляторных сетей. Сигналами в них выступают естественные метаболиты клетки, в частности индол, а также вторичные мессенджеры (p)ppGpp и цАМФ, взаимное влияние которых недостаточно изучено и представляет собой цель данного исследования. Родительский штамм *E. coli* BW25141 ((p)ppGpp⁺) и нокаут BW25141Δ*relA*Δ*spoT* ((p)ppGpp⁰), не способный синтезировать (p)ppGpp, культивировали на среде М9 с различным содержанием глюкозы (5.6 и 22.2 мМ), в присутствии триптофана в качестве субстрата для синтеза индола и в его отсутствие. Содержание глюкозы измеряли глюкозооксидазным методом, уровень индола – при помощи ВЭЖХ, концентрацию цАМФ – методом ELISA. Начало возрастания исходно низкого внутриклеточного содержания цАМФ совпадало с исчерпанием глюкозы в среде. Максимальный уровень накопления цАМФ в клетках был пропорционален концентрации исходно добавленной глюкозы. При этом (p)ppGpp⁰ мутант демонстрировал снижение максимального уровня цАМФ по сравнению с (p)ppGpp⁺ родителем, наиболее выраженное на среде с 22.2 мМ глюкозы. Таким образом, (p)ppGpp положительно регулирует образование цАМФ. Известно, что промотор триптофаназного оперона, ответственного за биосинтез индола, находится под положительным контролем механизма катабо-

литной репрессии. Поэтому в клетках (p)ppGpp⁺ штамма в условиях низкой скорости спонтанного образования индола на бестриптофановой среде значительно усиливался его синтез в ответ на возрастание уровня цАМФ при исчерпании глюкозы, чего не наблюдалось у (p)ppGpp⁰ мутанта с пониженным накоплением цАМФ. При добавке триптофана оба штамма демонстрировали высокую продукцию индола, что сопровождалось снижением уровня накопления цАМФ по сравнению с бестриптофановым контролем. Таким образом, (p)ppGpp при исчерпании глюкозы положительно регулирует накопление цАМФ и индола, который, в свою очередь, снижает образование цАМФ.

Ключевые слова: *Escherichia coli*; сигнальные молекулы; цАМФ; (p)ppGpp; индол; глюкоза; триптофан.

Introduction

The adaptation of bacteria to stress is based on a subtle tuning of regulatory networks by the changes in gene expression and enzyme activity. Furthermore, the metabolic products of bacterial cells, such as indole and second messengers (p)ppGpp and cAMP, are regulatory signals. The interaction of the metabolic pathways of these signaling molecules ensures the adaptation of bacteria to various changes in the environment, which is of great scientific interest.

Among the currently known nucleotide second messengers, cAMP (3',5'-cyclic adenosine monophosphate) and (p)ppGpp (guanosine tetra(penta)phosphate) were the first to be studied in most detail in the model organism *Escherichia coli*. They act as intracellular "secondary" signals, transducing information about the actual microenvironment of bacterial cells by transforming the extracellular "primary" signals into a cascade of intracellular physiological changes necessary to support important biological functions (Hengge et al., 2019).

cAMP is synthesized by the enzyme adenylate cyclase, encoded by the *cyoA* gene, that catalyzes the cyclization reaction of adenosine triphosphate (ATP) to form cAMP and inorganic pyrophosphate (Botsford, Harman, 1992) and is cleaved to form AMP by means of phosphodiesterases CpdA (Imamura et al., 1996) and DosP (Yoshimura-Suzuki et al., 2005). The global regulator cAMP plays an important role in many biological processes, including the regulation of growth, differentiation and general cell metabolism, division, starvation, anaerobiosis, carbon metabolism, stress reactions (Gosset et al., 2004), osmoregulation (Balsalobre et al., 2006), and quorum sensing (Zhou et al., 2008). Transcriptomic analysis has shown that more than 200 operons are under direct or indirect control of cAMP (Gutierrez-Ríos et al., 2007). cAMP in complex with the cAMP receptor protein (CRP) (Rickenberg, 1974) regulates the transcription of several hundred genes, some of which are involved in catabolic processes (Botsford, Harman, 1992), and thus participates in multiple regulatory networks.

Another well-known second messenger, (p)ppGpp, induces one of the main bacterial adaptive mechanisms – the stringent response (Potrykus, Cashel, 2008; Haurlyuk et al., 2015), affecting various aspects of bacterial physiology, such as persistence, virulence, biofilm formation, and others (Dalebroux et al., 2010; Maisonneuve, Gerdes, 2014). Since the (p)ppGpp alarmone is a global regulator that controls the expression of about 1,400 genes (Traxler et al., 2008), its main function is to regulate the growth and survival of bacterial cells in response to nutrient deficiency and various stress factors (Hengge et al., 2019). Elevated levels of (p)ppGpp delay bacterial growth by suppressing the production of RNA, DNA, and proteins (Mechold et al., 2013). In *E. coli* cells, (p)ppGpp levels are

determined by the balance between two alarmone synthetases RelA/SpoT, the homologues of (RSH) protein family – RelA with only synthesizing activity and the bifunctional enzyme SpoT, which exhibits both (p)ppGpp synthetase and hydrolyase activities. RelA synthetase is activated upon binding to ribosomes during amino acid starvation, when uncharged tRNAs bind to the A site of the ribosome. SpoT-dependent accumulation of (p)ppGpp in bacterial cells occurs when there is a lack of carbon sources, fatty acids, iron, nitrogen, phosphates, as well as under oxidative stress, etc. (Arenz et al., 2016). In *E. coli* cells, (p)ppGpp regulates gene transcription by changing the activity of RNA polymerase *via* direct interaction with the enzyme or indirectly, reducing the cellular pool of guanosine triphosphate (GTP) and, accordingly, the ATP/GTP ratio as a result of the consumption of GTP for (p)ppGpp synthesis (Dalebroux, Swanson, 2012; Haurlyuk et al., 2015).

Indole, a signaling molecule of stationary phase bacterial cells, also takes part in a signal transduction of changes in conditions of bacterial microenvironment. Indole production by *E. coli* cells, discovered in the early 20th century, has previously been used as a diagnostic marker to differentiate *E. coli* from other enteric bacteria. However, as it is now known, the ability of bacteria to synthesize indole is widespread among different species of microorganisms. It is produced by at least 27 genera of bacteria capable of synthesizing the enzyme tryptophanase TnaA, which breaks down tryptophan to produce indole, pyruvate and ammonium (Lee et al., 2007). Indole is imported into the cell mainly through Mtr permease (Yanofsky et al., 1991) and is exported from the cell using the multidrug efflux pump AcrEF (Kawamura-Sato et al., 1999). In *E. coli* cells, indole regulates such important cellular processes as biofilm formation (Lee et al., 2007), persistence (Kwan et al., 2015), the development of multidrug resistance (Hirakawa et al., 2005), changes in plasmid stability (Chant, Summers, 2007), motility (Bansal et al., 2007), and the ability to survive in mixed bacterial cultures (Chu et al., 2012).

Regulatory networks including the signaling molecules cAMP, (p)ppGpp and indole closely interact in bacterial cells to transduce intracellular or environmental signals for triggering a cascade of biochemical reactions leading to changes in cell gene expression profile and metabolism. One of the regulators of indole synthesis is cAMP, which, as part of the cAMP-CRP complex, positively regulates the transcription of tryptophanase operon responsible for indole production (Stewart, Yanofsky, 1985). We have previously shown that indole formation in *E. coli* is a (p)ppGpp-dependent process (Kashevarova et al., 2022). Decreased cAMP levels as a result of degradation by phosphodiesterases reduce indole production, increasing persistence (Kwan et al., 2015). At the same time, the works of other researchers showed that exogenous

addition of cAMP, with the participation of the cAMP-CRP regulatory complex controlling the expression of RelA synthetase, stimulated its expression and resulted in an increase in (p)ppGpp production and persistence. Moreover, the activity of the cAMP-CRP complex and the cAMP-CRP-controlled expression of RelA increased upon glucose depletion in both normal and persistent cells (Amato et al., 2013). Thus, regulatory signals cAMP and indole (Kwan et al., 2015), along with (p)ppGpp (Wood et al., 2013), are involved in multiple persistence pathways.

The purpose of our study is to study the effect of (p)ppGpp and indole on the formation of cAMP under conditions of depletion of the carbon substrate glucose.

Materials and methods

Strains and culture conditions. The objects of the study were the *E. coli* BW25141 parent strain (WT, (p)ppGpp⁺) (Datsenko, Wanner, 2000) and the BW25141Δ*relA*Δ*spoT* mutant ((p)ppGpp⁰) (laboratory collection). Cultures stored on Luria–Bertani (LB) agar (Sigma, USA) were cultivated for 5 hours at 37 °C in tubes with 5 mL of LB broth (Sigma, USA). Then, the cells were transferred to Erlenmeyer flasks with minimal M9 medium (50 mL) containing 5.6 mM (0.1 %) or 22.2 mM (0.4 %) glucose and cultivated at 120 rpm for 18 hours in a thermostated shaker GFL 1092 (GFL, Germany). Overnight cultures were diluted in 50 mL of M9 medium with appropriate concentrations of glucose to OD₆₀₀ = 0.2. Tryptophan (2 mM) (AppliChem, Germany) was added to some of the flasks, keeping the others as controls, and then cultivated under the same conditions for 168 hours. The optical density of the cultures was measured by the absorbance value at 600 nm on a UV-1650PC spectrophotometer (Shimadzu, Japan).

Gene engineering. The BW25141Δ*relA*Δ*spoT* deletion mutant was constructed using the FLP/FRT site-specific recombination system (Datsenko, Wanner, 2000).

Glucose content in the medium was determined by the glucose oxidase method using a Glucose-Vital kit (Vital Development Corporation, Russia). 5 μL of the samples were taken from the exponentially growing cultures with a time interval of one hour, and then once a day, and incubated with 1 mL of the kit reagent at room temperature for 15 minutes. When glucose is oxidized by glucose oxidase under the atmospheric pressure, an equimolar amount of hydrogen peroxide is formed that, in its turn, oxidizes chromogenic substrates by means of peroxidase to form a colored product proportionally to glucose concentration that was measured photometrically at 510 nm on the UV-1650PC spectrophotometer (Shimadzu, Japan).

Determination of indole. The concentration of indole in the medium was measured by high-performance liquid chromatography (Kim et al., 2013), with minor modifications, on an LC-20A chromatograph with an SPD-M20A spectrophotometric detector (Shimadzu, Japan), Luna C18 column (250×4.6 mm, 5 μm), SecurityGuard C18 pre-column (4×3 mm) (Phenomenex, USA). To determine extracellular indole, the samples were taken from the exponential phase cultures with a time interval of 2 hours and once a day from the stationary phase ones. Cells were pelleted by centrifugation. Supernatant samples in a volume of 20 μL were analyzed at a mobile phase flow rate of 1 mL/min (a mixture of acetonitrile (Kriochrome, Russia) and acetic acid in a 1:1 ratio) and +25 °C with detec-

tion at a wavelength of 280 nm. To calculate the indole content in the sample, the external standard method was used based on a previously constructed calibration curve.

Measurement of cAMP concentration. Quantitative determination of intracellular cAMP concentration was carried out by the competitive enzyme-linked immunosorbent assay (ELISA). 1.5 mL of the test culture was centrifuged at 12,000 rpm at +4 °C for 5 minutes. The pellet was resuspended in 300 μL of 0.1N HCl, kept for 5 minutes at +95 °C, followed by destruction of cells using a CPX-130 ultrasonic disintegrator (Cole-Parmer, USA), using a probe with a diameter of 6 mm (3 cycles for 30 seconds at an amplitude of 40 %). Cell residues were precipitated by centrifugation for 15 minutes. cAMP concentration was determined in the resulting supernatant after neutralization with 2 M Na₂CO₃ according to the manufacturer's instruction (RayBio® cAMP Enzyme Immunoassay). Experiments were carried out with at least two independent cultures. The data obtained were normalized to intracellular volume based on absolutely dry cell biomass.

Statistical analysis. The results obtained were processed statistically using the standard software package Statistica 6.0 (StatSoft Inc., USA). On the graphs, the means (at least three experiments) and the standard deviations are represented. The significance of differences was assessed using Student's *t*-test ($p \leq 0.05$).

Results

In the present work, the dynamics of glucose depletion and the formation of cAMP and indole in periodic cultures of the *E. coli* BW25141 wild-type strain and the Δ*relA*Δ*spoT* deletion mutant grown in M9 mineral medium with different contents of initially added glucose (5.6 and 22.2 mM) were studied. Overproduction of indole was induced by exogenous addition of the amino acid tryptophan (2 mM), a precursor of indole synthesis, in order to study its effect on the activity of metabolic processes, in particular, the consumption rate of carbon substrate (glucose) and cAMP formation.

When studying the glucose consumption in the strain periodic cultures, it was found that at an initially low concentration of glucose (5.6 mM), substrate depletion in the cells of both strains studied occurred in the first hours of cultivation (Fig. 1, *a*). Moreover, in the culture of the parental strain, glucose was completely consumed after 3 hours of growth, while (p)ppGpp absence in the cells of the Δ*relA*Δ*spoT* mutant led to a delay in substrate consumption by 2 hours as compared to the wild-type strain. At the same time, the presence of tryptophan in the medium had no effect in both cases.

In the culture of parent strain with high initial glucose concentration (22.2 mM), it was consumed for 12 hours after the cells were inoculated to a fresh nutrient medium (Fig. 1, *b*). The double Δ*relA*Δ*spoT* mutation, as well as at the low glucose concentration (5.6 mM), led to a slowing down of substrate utilization rate, so that glucose was consumed for more than 24 hours. However, the addition of tryptophan to the medium caused an acceleration of glucose consumption for more than 4 hours in the parental strain and for 24 hours in the (p)ppGpp⁰ mutant.

The onset of an increase in the initially low intracellular level of cAMP coincided with the depletion of glucose in the medium. Moreover, in cultures grown in the medium with

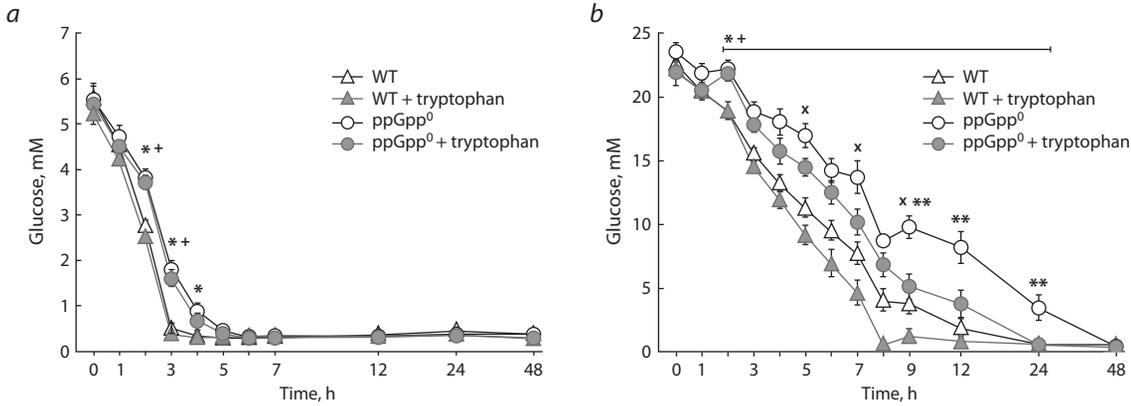


Fig. 1. Dynamics of glucose consumption in cultures of the parental strain (WT) and the (p)ppGpp⁰ mutant in the medium without tryptophan and with the addition of 2 mM tryptophan at 5.6 mM (a) and 22.2 mM glucose (b).

* Statistically significant difference between the (p)ppGpp⁰ mutant and the (p)ppGpp⁺ strain in the medium without tryptophan; + statistically significant difference between the (p)ppGpp⁰ mutant and the (p)ppGpp⁺ strain in the medium with the addition of 2 mM tryptophan; * statistically significant difference between the (p)ppGpp⁰ mutant and the (p)ppGpp⁺ strain in the tryptophan-free medium; ** statistically significant difference between the (p)ppGpp⁰ mutant in the medium with 2 mM tryptophan and the (p)ppGpp⁰ one in the tryptophan-free medium ($p \leq 0.05$).

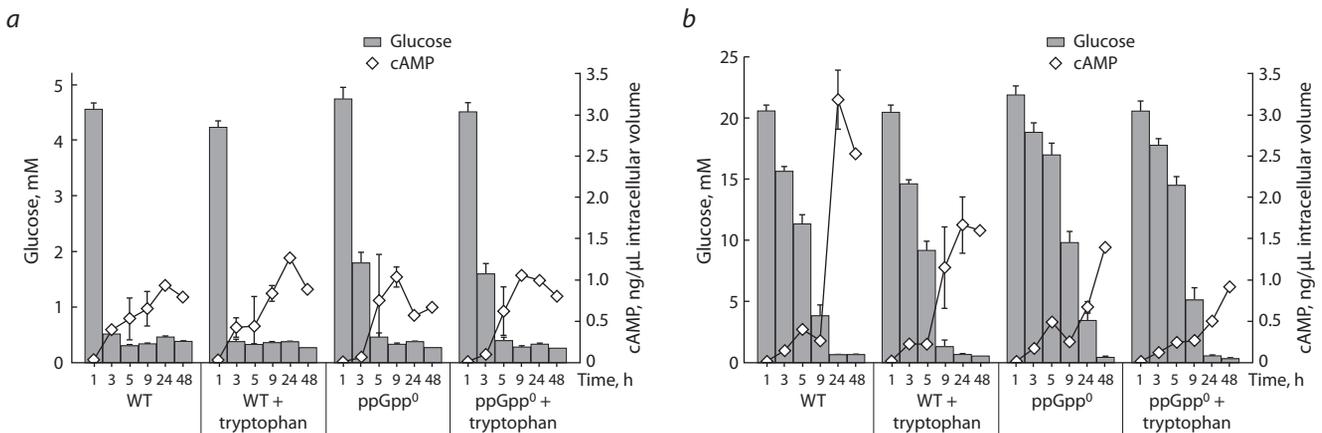


Fig. 2. Glucose consumption (histogram/columns) and cAMP formation (graph/diamonds) in cultures of the parental strain and the (p)ppGpp⁰ mutant in the medium without tryptophan and with the addition of 2 mM tryptophan at content of 5.6 mM (a) and 22.2 mM glucose (b).

5.6 mM glucose, the increase in cAMP concentration occurred earlier than in those grown in the medium with the addition of 22.2 mM glucose (Fig. 2). At low glucose content in the medium, the (p)ppGpp⁺ strain showed rapid depletion of glucose for the first 1–3 hours, which was accompanied by an active accumulation of cAMP during the same period (see Fig. 2, a). However, the double $\Delta relA \Delta spoT$ deletion led to a sharper increase in cAMP levels starting at 5 hours, which was not observed in the medium containing 22.2 mM glucose. By the end of the observed period of the first day of cultivation (9 hours of growth), higher values of cAMP accumulation were demonstrated by the mutant at 5.6 mM glucose, which was approximately 2 times higher than the cAMP level at 22.2 mM glucose.

In stationary cultures, a high concentration of initially added glucose (22.2 mM) in the medium led to higher levels of cAMP accumulation than at 5.6 mM glucose; however, in the case of the deletion mutant, these differences were less pronounced (Fig. 3).

The maximum values of cAMP accumulation in cells were proportional to the concentration of added glucose. At the same time, the (p)ppGpp⁰ knockout demonstrated a decrease in the maximum level of cAMP compared to the WT strain that was the most pronounced in the medium with 22.2 mM glucose. Maximum accumulation of cAMP at 5.6 mM glucose content in the mutant strain was achieved at 9 hours of cultivation with a subsequent decrease, whereas in the (p)ppGpp⁺ parent, by 72 hours in the tryptophan-free medium and 24 hours in the one supplemented with tryptophan (see Fig. 3, a). At high glucose content (22.2 mM), the maximum cAMP values in the WT strain were observed at 72 hours of growth both in the tryptophan-free medium and in the presence of the indole precursor (see Fig. 3, b). At the same time, the cAMP level in the (p)ppGpp⁰ knockout changed slightly throughout the entire cultivation period, especially in the medium supplemented with tryptophan, and was up to 6 times lower as compared to the WT strain. Thus, the absence of (p)ppGpp in the cells led to a decrease in cAMP production.

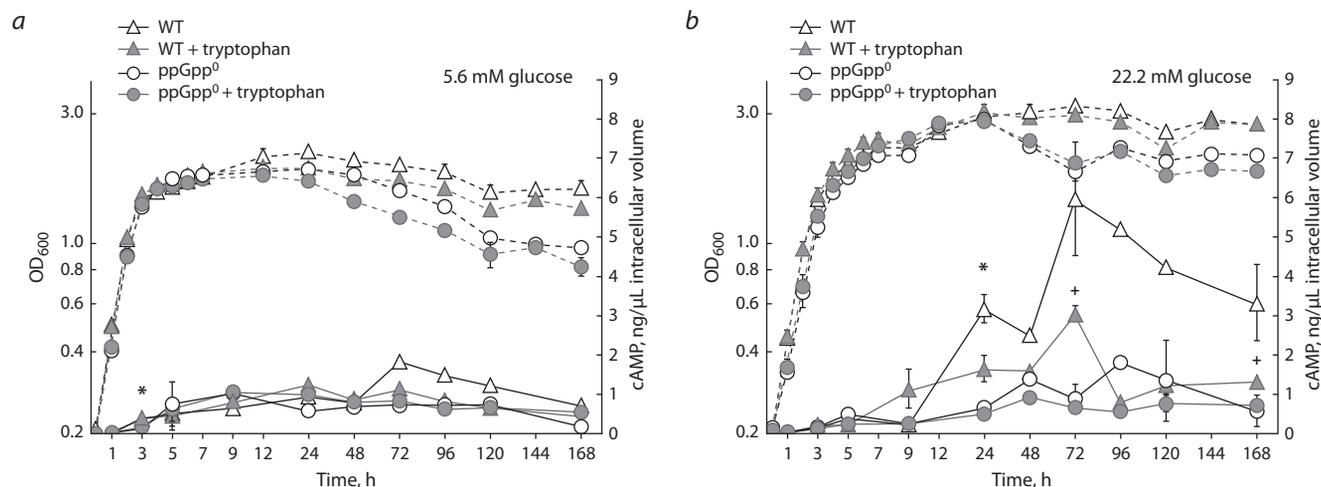


Fig. 3. Growth curves (dashed lines) and cAMP formation (solid lines) in the cultures of parent and mutant strains in the medium without tryptophan (open symbols) and with the addition of 2 mM tryptophan (filled symbols) at a content of 5.6 mM (a) and 22.2 mM glucose (b) in the medium.

* Statistically significant difference between the (p)ppGpp⁰ mutant and the (p)ppGpp⁺ strain in the medium without tryptophan; + statistically significant difference between the (p)ppGpp⁰ mutant and the (p)ppGpp⁺ strain at the supplementation of 2 mM tryptophan.

We have previously shown that the effect of (p)ppGpp – a stringent response signaling molecule – in the process of indole production regulation by *E. coli* cells depends on the glucose content in the cultural medium (Kashevarova et al., 2022). Therefore, (p)ppGpp positively regulates the synthesis of indole under the conditions of its spontaneous formation at high content of glucose in the tryptophan-free medium. However, this did not occur at low glucose concentration. At exogenous addition of 2 mM tryptophan, (p)ppGpp-dependent indole production was observed at both glucose concentrations studied, with a more pronounced effect at the high carbon substrate content. However, glucose limitation at the start of cultivation (in the medium with 5.6 mM glucose) led to an increase in the rate of indole accumulation and a reduction in the time spent to reach its maximum values (Fig. 4).

When studying the effect of indole overproduction on the synthesis of cAMP depending on the carbon substrate content, it was found that the addition of the amino acid tryptophan, which caused the accumulation of indole in the medium to 2.2–2.9 mM (see Fig. 4, c, d), did not have a significant effect on intracellular cAMP concentration in both parent and deletion strains at the studied glucose concentrations. However, there was an exception: at high glucose content, the (p)ppGpp⁺ strain demonstrated a decrease in the cAMP levels of up to 4 times compared to the tryptophan-free control in response to tryptophan addition (Fig. 5).

Discussion

Bacteria, exposed to various stress factors such as starvation, temperature, acid, oxidative stress, etc., have developed effective stress response mechanisms for cell adaptation to unfavorable conditions. Stress responses are induced by signaling molecules that trigger a chain of reactions leading to changes in gene expression and restructuring of cell metabolism, which allows bacteria to survive adverse effects.

In the present work, the mutual influence of the signaling molecules (p)ppGpp, cAMP, and indole upon depletion of

glucose in the culture medium was studied. Glucose is the preferred source of carbon and energy for *E. coli* cells, which have developed a complex regulatory network that coordinates gene expression, transport, and enzyme activity in response to the presence of this substrate (Gutierrez-Ríos et al., 2007). Carbon starvation induces a general stress response regulated by catabolic repression and the (p)ppGpp-mediated stringent response. In *E. coli*, genes controlled by the catabolic repression regulator protein (CRP) are RelA-dependent, indicating that (p)ppGpp is a global regulator of carbon starvation (Traxler et al., 2006) and demonstrating the interaction between catabolic repression and stringent response mechanisms. A recent study using fluorescence spectroscopy showed that (p)ppGpp is able to bind to the CRP protein with high affinity, and molecular docking results suggested that (p)ppGpp may negatively regulate the activity of this protein. Thus, CRP-controlled gene expression is modulated by (p)ppGpp as a result of its binding to CRP under starvation stress (Duysak et al., 2021).

In bacterial cells, intracellular levels of cAMP are dependent on many factors and, first of all, are determined by the ratio of the enzymes adenylate cyclase and phosphodiesterase activities. The cAMP signaling molecule is a part of the cAMP-CRP complex, which, on the one hand, is controlled by carbohydrates of the phosphoenolpyruvate phosphotransferase system (Deutscher et al., 2006), and, on the other, regulates the catabolite repression of carbon (Botsford, Harman, 1992). The presence of glucose in the growth medium suppresses the expression of enzymes that catalyze the metabolism of other carbon sources, reducing the levels of both CRP protein and cAMP (Ishizuka et al., 1994). In *E. coli*, cellular cAMP levels are inversely proportional to the concentration of carbon source, resulting in low cAMP content in the presence of an easily metabolized carbon source (glucose). In addition to carbon catabolism, cAMP-CRP regulon controls important cellular functions associated with stress response, cell division, and amino acid metabolism, including the *tnaA* (tryptophanase) gene, which is responsible for the production

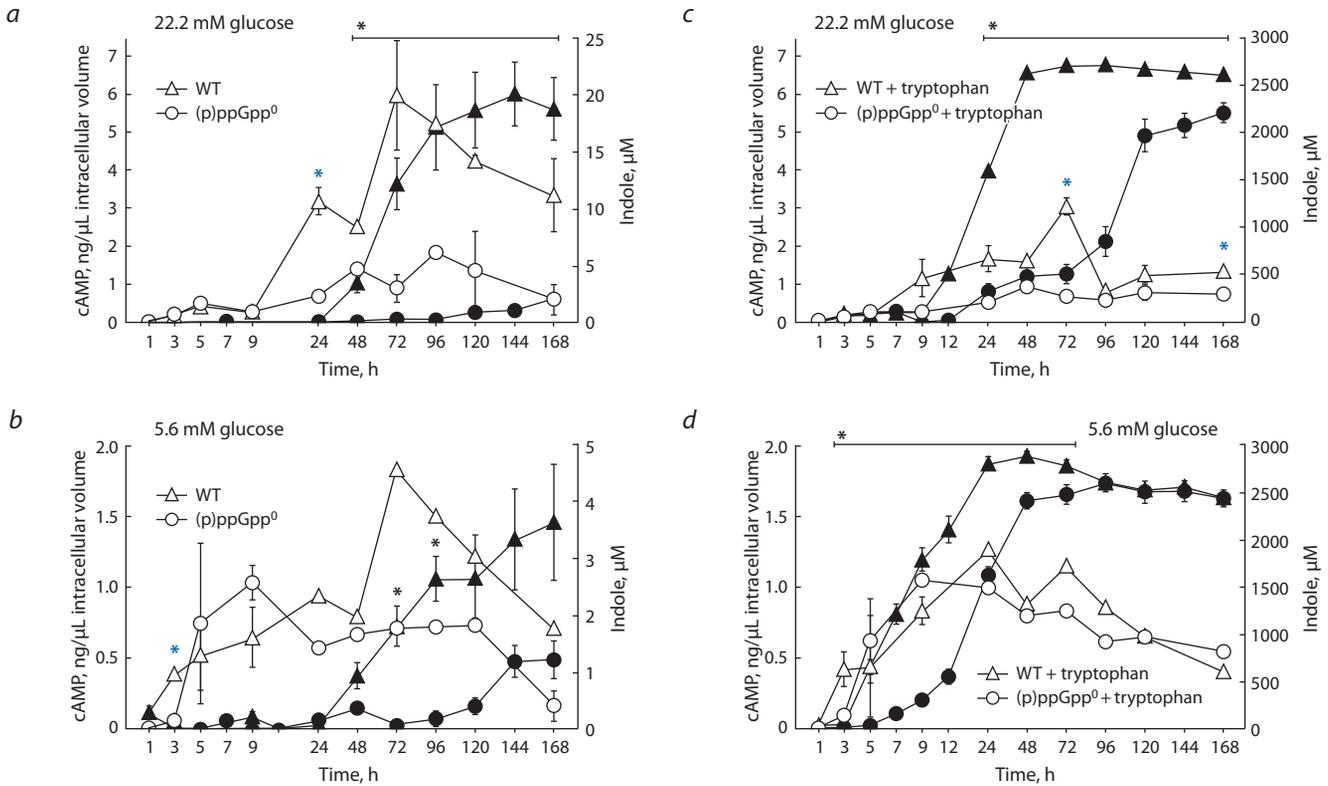


Fig. 4. Indole production (filled symbols) and cAMP formation (open symbols) in cultures of the parent strain (WT) and the (p)ppGpp⁰ mutant in the medium without tryptophan (a, b) and with the addition of 2 mM tryptophan (c, d) at 22.2 mM (a, c) and 5.6 mM glucose (b, d).

* Statistically significant difference between the (p)ppGpp⁰ mutant and the (p)ppGpp⁺ strain in terms of indole and cAMP production (black and blue, respectively) ($p \leq 0.05$).

of one of the *E. coli* metabolites, indole, from the amino acid tryptophan (Isaacs et al., 1994).

In this work, the dependence of the dynamics of cAMP and indole synthesis on the presence of (p)ppGpp and glucose contents in the growth medium has been studied. In the experiments, two glucose concentrations of 5.6 and 22.2 mM have been used, which we have designated as low and high glucose contents, respectively. When glucose is depleted in the medium with tryptophan, the metabolism of bacterial cells switches to the utilization of this substrate, which *E. coli* cells are able to use as the only source of carbon and nitrogen due to the action of the inducible enzyme tryptophanase, which breaks down tryptophan to form indole (Yanofsky et al., 1991).

Our studies have shown that the absence of (p)ppGpp in cells reduced their ability to utilize glucose, which was the most pronounced at high glucose content in the growth medium (22.2 mM). However, an acceleration of glucose consumption from the medium supplemented with tryptophan was demonstrated (see Fig. 1). The ability of bacterial cells to produce (p)ppGpp plays an important role in glucose metabolism, which has been shown for both Gram-positive and Gram-negative bacteria (Oh et al., 2015). In the process of bacterial adaptation to glucose starvation, in addition to stringent response, in which cell growth and the synthesis of macromolecules are inhibited, an expanded adaptive response is formed, including inhibition of glycolysis and metabolic transitions mediated through the mechanism of catabolite repression (Zhang et al., 2016). Our experiments have shown

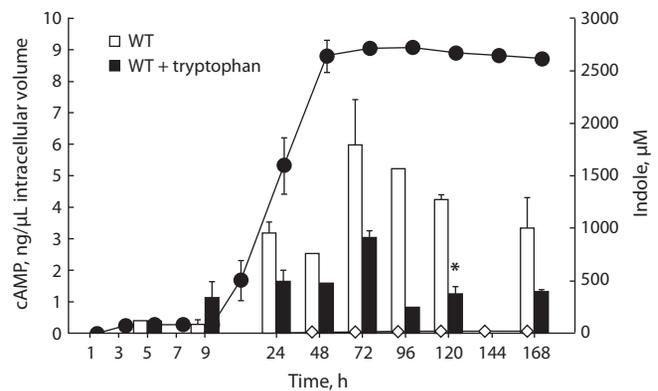


Fig. 5. Indole production (graph) and cAMP formation (histogram) in the parent strain in the medium without tryptophan (diamonds/white columns) and with the addition of 2 mM tryptophan (circles/black columns) at 22.2 mM glucose.

* Statistically significant difference between the (p)ppGpp⁺ strain supplemented with 2 mM tryptophan and the same one grown in the tryptophan-free medium ($p \leq 0.05$).

that the depletion of glucose in the medium is accompanied by an increase in the initially low intracellular cAMP content. The maximum level of cAMP accumulation in bacterial cells has been shown to be proportional to the concentration of the initially added substrate. At low glucose concentration (5.6 mM), an increase in the cAMP level occurred earlier

than at 22.2 mM glucose (see Fig. 2). At 5.6 mM glucose, the maximum values of intracellular cAMP concentration over a 168-hour period of observation were achieved already at the first day of cultivation in the (p)ppGpp⁰ mutant and at 24 hours in the parental strain with the addition of tryptophan. At the same time, in the tryptophan-free culture of the (p)ppGpp⁺ strain, the maximum cAMP level was observed at 72 hours of cultivation (see Fig. 3, a), as well as under the conditions of 22.2 mM glucose addition to the medium (see Fig. 3, b). At initially high glucose content, the level of cAMP was higher in the parental strain as compared to the deletion mutant both in the tryptophan-free medium and in its presence (see Fig. 3, b). Thus, the absence of (p)ppGpp in cells led to a decrease in cAMP formation. These data allow us to assume that the process of cAMP synthesis in *E. coli* cells is positively regulated by (p)ppGpp.

The second messenger (p)ppGpp is able to modulate gene expression in response to various stress factors in different bacterial species (Irving, Corrigan, 2018). It is a key regulator of stringent response, one of the most important protective mechanisms in bacterial adaptation induced primarily by starvation stress (Traxler et al., 2006, 2008). In Gram-negative bacteria, (p)ppGpp regulates transcription initiation by directly binding to RNA polymerase. In this case, the stimulatory or inhibitory effects of (p)ppGpp on transcription depend on the composition of the discriminatory sequences (AT- or GC-rich) (Sanchez-Vazquez et al., 2019). In addition, (p)ppGpp can directly bind to proteins and thereby change their catalytic activity. Thus, in the work (Ro et al., 2021), it was shown that (p)ppGpp promotes acetylation of the CRP protein, which, as a part of the cAMP-CRP complex, plays a key role in the regulation of gene expression in *E. coli*. However, there are no significant differences between the parent strain and the mutant one with a deletion of one of two *E. coli* *ΔrelA* alarmone synthetases in either the levels of cAMP content or the expression of adenylate cyclase *cyaA* responsible for the production of cAMP.

Our results have demonstrated the effect of (p)ppGpp on the level of cAMP content, which is displayed in an increase in the concentration of cAMP in the (p)ppGpp⁺ strain as compared to the mutant strain with the double *ΔrelAΔspoT* deletion by 2.8 and 6 times at 5.6 and 22.2 mM glucose content, respectively, at 72 hours of growth in the tryptophan-free medium (see Fig. 3). A similar effect of (p)ppGpp on the level of cAMP was also observed under excessive accumulation of indole, but only at high initial glucose concentration in the medium (22.2 mM) (see Fig. 3, b). However, the difference in cAMP content between the WT strain and the knockout one disappeared under glucose limitation (5.6 mM, see Fig. 3, a). The lack of (p)ppGpp effect on cAMP levels (Ro et al., 2021) could probably be explained by the activity of SpoT synthetase leading to the accumulation of (p)ppGpp, while in our experiments the complete absence of stringent response signaling molecule should be ensured by the double *ΔrelAΔspoT* deletion. Interestingly, the presence of (p)ppGpp in cells accelerated the onset of cAMP synthesis under the conditions of glucose limitation (at the initial content of 5.6 mM glucose in the growth medium), which was observed after 3 hours of growth by a statistically significant 4-fold increase in intracellular cAMP concentration in the (p)ppGpp⁺ strain as compared to the *ΔrelAΔspoT*

mutant in the tryptophan-free medium. However, at the next two sampling points – 5 and 9 hours of cultivation – the (p)ppGpp⁰ knockout was superior compared to the parent strain in the level of cAMP production (see Fig. 4, b), which was also displayed under tryptophan supplementation (see Fig. 4, d).

The present study shows that (p)ppGpp is involved in the regulation of cAMP production, as well as in indole biosynthesis. It is known that the *tnaCAB* operon responsible for indole biosynthesis is positively regulated by the cAMP-CRP-dependent mechanism of catabolite repression and is induced at the transcriptional level upon depletion of carbohydrates and cell transition to the stationary phase (Stewart, Yanofsky, 1985). This explains the observed increase in the cAMP content to levels of 1.8 and 6 ng/μL during cell growth in the tryptophan-free medium at 5.6 and 22.2 mM glucose, respectively (see Fig. 3). This was accompanied by a significant increase in indole formation of up to 10 times at high glucose content in the medium, which is typical only for the parent strain, but not for the mutant one (see Fig. 4, a). Based on these data, we concluded that the alarmone (p)ppGpp along with catabolite repression can positively regulate the process of indole formation. As it follows from Fig. 4, a, b, the intracellular concentration of cAMP was decreased by 6 and 3 times at 22.2 and 5.6 mM glucose, respectively, in the *ΔrelAΔspoT* mutant as compared to the (p)ppGpp⁺ strain under growth in the tryptophan-free medium. This indicates a positive role of (p)ppGpp in the regulation of both cAMP synthesis and indole production. Under the addition of tryptophan, both the parent and the mutant strains have demonstrated high indole production at both glucose concentrations studied (see Fig. 4, c, d). The increase in indole formation occurred earlier in the medium with low glucose content compared to the one with 22.2 mM glucose, which may be due to the earlier formation of cAMP in the cells. At the same time, the mutant has also demonstrated a time delay in indole accumulation, which was more pronounced at high initial glucose content in the medium (see Fig. 4, c). Thus, in *E. coli* cells, indole production has been shown to be under the positive control of the second messengers (p)ppGpp and cAMP under glucose depletion during batch culture growth.

As shown by our experiments, the cAMP level in *E. coli* cells depends to some extent on the indole content in the medium. The presence of tryptophan, a substrate for indole synthesis, led to overproduction of indole in the cultures of both the WT strain and the mutant. However, a decrease in cAMP concentration was observed only in the parent strain, which was more pronounced at 22.2 mM glucose (see Fig. 5). However, high indole production in knockout cell culture did not produce a significant effect on the cAMP level.

Therefore, the results obtained in this work demonstrate that the absence of the alarmone (p)ppGpp in *E. coli* cells led to a decrease in cAMP formation and deceleration in glucose metabolism. An increase in the level of cAMP in response to glucose depletion was accompanied by the synthesis of indole with a pronounced decrease in its accumulation in the (p)ppGpp⁰ mutant as compared to the parent strain. This was observed in both the tryptophan-free medium and at the tryptophan addition. In cells of the (p)ppGpp⁺ strain growing in the tryptophan-free medium at a low rate of spontaneous

indole formation, its synthesis was significantly increased in response to an increase in the level of cAMP upon depletion of glucose. This phenomenon was not observed in the (p)ppGpp⁰ mutant that was characterized by a reduced accumulation of cAMP. These data indicate a positive role of (p)ppGpp in the regulation of indole formation, along with the cAMP-mediated catabolic repression. At the addition of tryptophan, both the parental and (p)ppGpp⁰ strains have demonstrated high indole production, which was accompanied by a decrease in the level of cAMP accumulation in comparison with the tryptophan-free control and was the most pronounced in the (p)ppGpp⁺ strain at 22.2 mM glucose. Thus, at glucose depletion, the alarmone (p)ppGpp positively regulates the accumulation of cAMP and indole, and the latter, in its turn, reduces the formation of cAMP.

Conclusion

The survival of bacteria in continuously changing environmental conditions is due to a complex, formed during the evolutionary process, system of numerous interacting regulatory networks that control the diverse physiological functions of bacteria under various stress factors. This work studies the interaction of signaling molecules cAMP, (p)ppGpp, and indole, under the conditions of substrate limitation. The results obtained demonstrate that (p)ppGpp functions as a positive regulator of such processes as glucose metabolism, cAMP synthesis, and indole production. The intracellular level of cAMP accumulation depends on both glucose availability in the medium and the stringent response alarmone (p)ppGpp. The absence of (p)ppGpp in cells reduces their ability to produce indole under the conditions of a low rate of its spontaneous formation during cell growth in the tryptophan-free medium and decelerates the rate of indole accumulation in the medium supplemented with tryptophan. Indole biosynthesis in *E. coli* is positively regulated by the signaling molecules (p)ppGpp and cAMP at substrate limitation. This gives us reason to conclude that (p)ppGpp-dependent indole production is mediated through changes in the level of cAMP in cells. The results obtained demonstrate the modulatory effect of (p)ppGpp on the expression of genes of the tryptophanase operon regulated by means of the cAMP-CRP-dependent mechanism of catabolite repression, the removal of which is induced in response to glucose depletion.

References

- Amato S.M., Orman M.A., Brynildsen M.P. Metabolic control of persister formation in *Escherichia coli*. *Mol. Cell*. 2013;50(4):475-487. DOI 10.1016/j.molcel.2013.04.002
- Arenz S., Abdelshahid M., Sohmen D., Payoe R., Starosta A.L., Berninghausen O., Haurlyuk V., Beckmann R., Wilson D.N. The stringent factor RelA adopts an open conformation on the ribosome to stimulate ppGpp synthesis. *Nucleic Acids Res*. 2016;44(13):6471-6481. DOI 10.1093/nar/gkw470
- Balsalobre C., Johansson J., Uhlin B.E. Cyclic AMP-dependent osmoregulation of *crp* gene expression in *Escherichia coli*. *J. Bacteriol*. 2006;188(16):5935-5944. DOI 10.1128/JB.00235-06
- Bansal T., Englert D., Lee J., Hegde M., Wood T.K., Jayaraman A. Differential effects of epinephrine, norepinephrine, and indole on *Escherichia coli* O157: H7 chemotaxis, colonization, and gene expression. *Infect. Immun*. 2007;75(9):4597-4607. DOI 10.1128/IAI.00630-07
- Botsford J.L., Harman J.G. Cyclic AMP in prokaryotes. *Microbiol. Rev*. 1992;56(1):100-122. DOI 10.1128/mr.56.1.100-122.1992
- Chant E.L., Summers D.K. Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. *Mol. Microbiol*. 2007;63(1):35-43. DOI 10.1111/j.1365-2958.2006.05481.x
- Chu W., Zere T.R., Weber M.M., Wood T.K., Whiteley M., Hidalgo-Romano B., Valenzuela E.J., McLean R.J.C. Indole production promotes *Escherichia coli* mixed-culture growth with *Pseudomonas aeruginosa* by inhibiting quorum signaling. *Appl. Environ. Microbiol*. 2012;78(2):411-419. DOI 10.1128/AEM.06396-11
- Dalebroux Z.D., Svensson S.L., Gaynor E.C., Swanson M.S. ppGpp conjures bacterial virulence. *Microbiol. Mol. Biol. Rev*. 2010;74(2):171-199. DOI 10.1128/MMBR.00046-09
- Dalebroux Z.D., Swanson M.S. ppGpp: magic beyond RNA polymerase. *Nat. Rev. Microbiol*. 2012;10(3):203-212. DOI 10.1038/nrmicro2720
- Datsenko K.A., Wanner B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA*. 2000;97(12):6640-6645. DOI 10.1073/pnas.120163297
- Deutscher J., Francke C., Postma P.W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev*. 2006;70(4):939-1031. DOI 10.1128/MMBR.00024-06
- Duysak T., Tran T.T., Afzal A.R., Jung C. Fluorescence spectroscopic analysis of ppGpp binding to cAMP receptor protein and histone-like nucleoid structuring protein. *Int. J. Mol. Sci*. 2021;22(15):7871. DOI 10.3390/ijms22157871
- Gosset G., Zhang Z., Nayyar S., Cuevas W.A., Saier M.H. Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. *J. Bacteriol*. 2004;186(11):3516-3524. DOI 10.1128/JB.186.11.3516-3524.2004
- Gutierrez-Ríos R.M., Freyre-Gonzalez J.A., Resendis O., Collado-Vides J., Saier M., Gosset G. Identification of regulatory network topological units coordinating the genome-wide transcriptional response to glucose in *Escherichia coli*. *BMC Microbiol*. 2007;7:53. DOI 10.1186/1471-2180-7-53
- Haurlyuk V., Atkinson G.S., Murakami K.S., Tenson T., Gerdes K. Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat. Rev. Microbiol*. 2015;13(5):298-309. DOI 10.1038/nrmicro3448
- Hengge R., Häussler S., Pruteanu M., Stülke J., Tschowri N., Turgay K. Recent advances and current trends in nucleotide second messenger signaling in bacteria. *J. Mol. Biol*. 2019;431(5):908-927. DOI 10.1016/j.jmb.2019.01.014
- Hirakawa H., Inazumi Y., Masaki T., Hirata T., Yamaguchi A. Indole induces the expression of multidrug exporter genes in *Escherichia coli*. *Mol. Microbiol*. 2005;55(4):1113-1126. DOI 10.1111/j.1365-2958.2004.04449
- Imamura R., Yamanaka K., Ogura T., Hiraga S., Fujita N., Ishihama A., Niki H. Identification of the *cpdA* gene encoding cyclic 3',5'-adenosine monophosphate phosphodiesterase in *Escherichia coli*. *J. Biol. Chem*. 1996;271(41):25423-25429. DOI 10.1074/jbc.271.41.25423
- Irving S.E., Corrigan R.M. Triggering the stringent response: signals responsible for activating (p)ppGpp synthesis in bacteria. *Microbiology*. 2018;164(3):268-276. DOI 10.1099/mic.0.000621
- Isaacs H., Chao D., Yanofsky C., Saier M.H. Mechanism of catabolite repression of tryptophanase synthesis in *Escherichia coli*. *Microbiology*. 1994;140(8):2125-2134. DOI 10.1099/13500872-140-8-2125
- Ishizuka H., Hanamura A., Inada T., Aiba H. Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the *crp* gene. *EMBO J*. 1994;13(13):3077-3082. DOI 10.1002/j.1460-2075.1994.tb06606.x
- Kashevarova N.M., Akhova A.V., Khaova E.A., Tkachenko A.G. Role of alarmone (p)ppGpp in the regulation of indole formation depending on glucose content in *Escherichia coli*. *Acta Biomedica Scien*

- tifica*. 2022;7(3):162-168. DOI 10.29413/ABS.2022-7.3.17 (in Russian)
- Kawamura-Sato K., Shibayama K., Horii T., Iimuma Y., Arakawa Y., Ohta M. Role of multiple efflux pumps in *Escherichia coli* in indole expulsion. *FEMS Microbiol. Lett.* 1999;179(2):345-352. DOI 10.1016/S0378-1097(99)00433-4
- Kim D., Sitepu I.R., Hashidokoa Y. Induction of biofilm formation in the betaproteobacterium *Burkholderia unamae* CK43B exposed to exogenous indole and gallic acid. *Appl. Environ. Microbiol.* 2013; 79(16):4845-4852. DOI 10.1128/AEM.01209-13
- Kwan B.W., Osbourne D.O., Hu Y., Benedik M.J., Wood T.K. Phosphodiesterase DosP increases persistence by reducing cAMP which reduces the signal indole. *Biotechnol. Bioeng.* 2015;112(3):588-600. DOI 10.1002/bit.25456
- Lee J., Jayaraman A., Wood T.K. Indole is an inter-species biofilm signal mediated by SdiA. *BMC Microbiol.* 2007;7:42. DOI 10.1186/1471-2180-7-42
- Maisonneuve E., Gerdes K. Molecular mechanisms underlying bacterial persisters. *Cell.* 2014;157(3):539-548. DOI 10.1016/j.cell.2014.02.050
- Mechold U., Potrykus K., Murphy H., Murakami K.S., Cashel M. Differential regulation by ppGpp versus pppGpp in *Escherichia coli*. *Nucleic Acids Res.* 2013;41(12):6175-6189. DOI 10.1093/nar/gkt302
- Oh Y.T., Lee K., Bari W., Raskin D.M., Yoon S.S. (p)ppGpp, a small nucleotide regulator, directs the metabolic fate of glucose in *Vibrio cholera*. *J. Biol. Chem.* 2015;290(21):13178-13190. DOI 10.1074/jbc.M115.640466
- Potrykus K., Cashel M. (p)ppGpp: still magical? *Annu. Rev. Microbiol.* 2008;62:35-51. DOI 10.1146/annurev.micro.62.081307.162903
- Rickenberg H.V. Cyclic AMP in prokaryotes. *Annu. Rev. Microbiol.* 1974;28:353-369. DOI 10.1146/annurev.mi.28.100174.002033
- Ro C., Cashel M., Fernández-Col L. The secondary messenger ppGpp interferes with cAMP-CRP regulon by promoting CRP acetylation in *Escherichia coli*. *PLoS One.* 2021;16(10):e0259067. DOI 10.1371/journal.pone.0259067
- Sanchez-Vazquez P., Dewey C.N., Kitten N., Ross W., Gourse R.L. Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proc. Natl. Acad. Sci. USA.* 2019;116(17):8310-8319. DOI 10.1073/pnas.1819682116
- Stewart V., Yanofsky C. Evidence for transcription antitermination control of tryptophanase operon expression in *Escherichia coli* K-12. *J. Bacteriol.* 1985;164(2):731-740. DOI 10.1128/jb.164.2.731-740.1985
- Traxler M.F., Chang D.E., Conway T. Guanosine 3',5'-bispyrophosphate coordinates global gene expression during glucose-lactose diauxie in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* 2006;103(7):2374-2379. DOI 10.1073/pnas.0510995103
- Traxler M.F., Summers S.M., Nguyen H., Zacharia V.M., Smith J.T., Conway T. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol. Microbiol.* 2008;68(5):1128-1148. DOI 10.1111/j.1365-2958.2008.06229.x
- Wood T.K., Knabel S.J., Kwan B.W. Bacterial persister cell formation and dormancy. *Appl. Environ. Microbiol.* 2013;79(23):7116-7121. DOI 10.1128/AEM.02636-13
- Yanofsky C., Horn V., Gollnick P. Physiological studies of tryptophan transport and tryptophanase operon induction in *Escherichia coli*. *J. Bacteriol.* 1991;173(19):6009-6017. DOI 10.1128/jb.173.19.6009-6017.1991
- Yoshimura-Suzuki T., Sagami I., Yokota N., Kurokawa H., Shimizu T. DOS(Ec), a heme-regulated phosphodiesterase, plays an important role in the regulation of the cyclic AMP level in *Escherichia coli*. *J. Bacteriol.* 2005;187(19):6678-6682. DOI 10.1128/JB.187.19.6678-6682.2005
- Zhang T., Zhu J., Wei S., Luo Q., Li L., Li S., Tucker A., Shao H., Zhou R. The roles of RelA/(p)ppGpp in glucose-starvation induced adaptive response in the zoonotic *Streptococcus suis*. *Sci. Rep.* 2016;6:27169. DOI 10.1038/srep27169
- Zhou X., Meng X., Sun B. An EAL domain protein and cyclic AMP contribute to the interaction between the two quorum sensing systems in *Escherichia coli*. *Cell Res.* 2008;18(9):937-948. DOI 10.1038/cr.2008.67

ORCID

N.M. Kashevarova orcid.org/0000-0003-3751-8156
E.A. Khaova orcid.org/0000-0003-4457-2652
A.G. Tkachenko orcid.org/0000-0002-8631-8583

Acknowledgements. The research was supported by the Ministry of Science and Higher Education of the Russian Federation (124020500028-4).

Conflict of interest. The authors declare no conflict of interest.

Received July 14, 2023. Revised November 13, 2023. Accepted November 15, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Effects of polyamines and indole on the expression of ribosome hibernation factors in *Escherichia coli* at the translational level

E.A. Khaova , A.G. Tkachenko

Institute of Ecology and Genetics of Microorganisms of the Ural Branch of the Russian Academy of Sciences,
Perm Federal Research Center of the Ural Branch of the Russian Academy of Sciences, Perm, Russia
 akkuzina-elena510@mail.ru

Abstract. Polyamines and indole are small regulatory molecules that are involved in the adaptation to stress in bacteria, including the regulation of gene expression. Genes, the translation of which is under the regulatory effects of polyamines, form the polyamine modulon. Previously, we showed that polyamines upregulated the transcription of genes encoding the ribosome hibernation factors RMF, RaiA, SRA, EttA and RsfS in *Escherichia coli*. At the same time, indole affected the expression at the transcriptional level of only the *raiA* and *rmf* genes. Ribosome hibernation factors reversibly inhibit translation under stress conditions, including exposure to antibiotics, to avoid resource waste and to conserve ribosomes for a quick restoration of their functions when favorable conditions occur. In this work, we have studied the influence of indole on the expression of the *raiA* and *rmf* genes at the translational level and regulatory effects of the polyamines putrescine, cadaverine and spermidine on the translation of the *rmf*, *raiA*, *sra*, *ettA* and *rsfS* genes. We have analyzed the mRNA primary structures of the studied genes and the predicted mRNA secondary structures obtained by using the RNAfold program for the availability of polyamine modulon features. We have found that all of the studied genes contain specific features typical of the polyamine modulon. Furthermore, to investigate the influence of polyamines and indole on the translation of the studied genes, we have constructed the translational reporter *lacZ*-fusions by using the pRS552/λRS45 system. According to the results obtained, polyamines upregulated the expression of the *rmf*, *raiA* and *sra* genes, the highest expression of which was observed at the stationary phase, but did not affect the translation of the *ettA* and *rsfS* genes, the highest expression of which took place during the exponential phase. The stimulatory effects were polyamine-specific and observed at the stationary phase, when bacteria are under multiple stresses. In addition, the data obtained demonstrated that indole significantly inhibited translation of the *raiA* and *rmf* genes, despite the stimulatory effect on their transcription. This can suggest the activity of a posttranscriptional regulatory mechanism of indole on gene expression.

Key words: polyamines; polyamine modulon; indole; ribosome hibernation factors; reporter gene fusions; gene expression; adaptation to stress.

For citation: Khaova E.A., Tkachenko A.G. Effects of polyamines and indole on the expression of ribosome hibernation factors in *Escherichia coli* at the translational level. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(1):24-32. DOI 10.18699/vjgb-24-04

Регуляторные эффекты полиаминов и индола на экспрессию факторов гибернации рибосом у *Escherichia coli* на уровне трансляции

Е.А. Хаова , А.Г. Ткаченко

Институт экологии и генетики микроорганизмов Уральского отделения Российской академии наук –
филиал Пермского федерального исследовательского центра Уральского отделения Российской академии наук, Пермь, Россия
 akkuzina-elena510@mail.ru

Аннотация. Полиамины и индол – регуляторные молекулы, которые участвуют в адаптации к стрессу у бактерий, включая регуляцию генной экспрессии. Гены, трансляция которых находится под регуляторным влиянием полиаминов, составляют полиаминовый модулон. Ранее нами показано, что полиамины стимулируют транскрипцию генов, кодирующих факторы гибернации рибосом RMF, RaiA, SRA, EttA, RsfS у *Escherichia coli*, а эффект индола ограничивался лишь двумя из них – *raiA* и *rmf*. Факторы гибернации рибосом обратимо ингибируют трансляцию в условиях стресса с целью экономии клеточных ресурсов, играя ключевую роль в выживании бактерий, в том числе при воздействии антибиотиков. Данная работа посвящена изучению влияния индола на экспрессию генов *raiA* и *rmf* на трансляционном уровне, а также регуляторных эффектов полиаминов путресцина, кадаверина и спермидина на трансляцию генов *rmf*, *raiA*, *sra*, *ettA*, *rsfS*. Проанализи-

ровая первичная структура мРНК, а также полученные с помощью программы RNAfold модели вторичных структур мРНК, мы установили, что все исследуемые гены имеют специфические признаки полиаминового модулона. Для изучения влияния полиаминов и индола на трансляцию исследуемых генов были сконструированы трансляционные репортерные *lacZ*-слияния с использованием pRS552/ARS45 системы. Согласно полученным результатам, полиамины стимулируют экспрессию «стационарно-фазных» генов *rmf*, *raiA*, *sra*, но не оказывают влияния на трансляцию генов *ettA* и *rsfS*, наибольшая экспрессия которых наблюдается в экспоненциальной фазе роста. Стимулирующий эффект специфичен для различных полиаминов и во всех случаях наблюдается в стационарной фазе, когда клетки подвержены множественному стрессорному воздействию. Кроме того, полученные данные демонстрируют значительный ингибирующий эффект индола на экспрессию *raiA* и *rmf* на уровне трансляции, несмотря на его стимулирующее действие на транскрипцию данных генов, что может являться признаком функционирования посттранскрипционного механизма их регуляции. Ключевые слова: полиамины; полиаминовый модулон; индол; факторы гибернации рибосом; репортерные генные слияния; генная экспрессия; адаптация к стрессу.

Introduction

Being the normal metabolites of bacteria, polyamines and indole are involved in a variety of cellular processes, including adaptation to stress, antibiotic resistance, biofilm formation, quorum sensing and persistence (Rhee et al., 2007; Shah, Swiatlo, 2008; Tkachenko et al., 2012, 2014; Gaimster et al., 2014; Lee et al., 2015; Miller-Fleming et al., 2015; Michael, 2018; Kim et al., 2020; Zarkan et al., 2020; Lang et al., 2021). Biogenic polyamines are aliphatic polycations synthesized from amino acids and present in almost all biological materials. Bacteria are able to produce predominantly putrescine, cadaverine and spermidine (Tabor C.W., Tabor H., 1985; Michael, 2016). In turn, indole is a heterocyclic aromatic compound, which is produced from tryptophan by many bacterial species and is involved in interspecies and interkingdom signaling (Zarkan et al., 2020).

One of the ways to realize the regulatory effects of these metabolites is the modulation of gene expression (Igarashi, Kashiwagi, 2006, 2018; Kusano et al., 2008; Shah, Swiatlo, 2008; Miller-Fleming et al., 2015; Zarkan et al., 2020; Lang et al., 2021). Intracellular polyamines are mainly presented by complexes with RNA, including mRNA. Genes, the expression of which is upregulated by polyamines at the translational level, form the polyamine modulon (Igarashi, Kashiwagi, 2006, 2018). At the same time, there are data indicating that the regulatory activity of both polyamines and indole can be displayed on different levels of gene expression (Miller-Fleming et al., 2015; Lang et al., 2021; Khaova et al., 2022). Signaling molecules are able to form the regulatory networks that can form responses to various stresses (Tkachenko, 2012). Recently, data were produced about the mutual influence of polyamines and indole on *Escherichia coli* metabolism. In particular, exogenous indole was able to increase the intracellular content of putrescine and spermidine, whereas the addition of spermine, which is the predominant product of eukaryotes, was capable of increasing the indole content in the cultural medium (Nesterova et al., 2021). The functioning of regulatory networks is aimed at optimizing responses of bacterial cells to changes in environmental conditions (Tkachenko, 2012). Polyamines and indole are known to have many different effects on cellular processes, but their molecular targets and mechanisms of action are still not fully understood (Rhee et al., 2007; Kusano et al., 2008; Shah, Swiatlo, 2008; Lee et

al., 2015; Miller-Fleming et al., 2015; Michael, 2018; Zarkan et al., 2020).

Previously, we studied the influence of polyamines and indole on transcription of the *rmf*, *raiA*, *sra*, *ettA*, *rsfS* genes, encoding ribosome hibernation factors, in *E. coli* (Khaova et al., 2022). These factors are able to reversibly inhibit ribosomes under the conditions of nutrient depletion and other stresses in order to save cell resources and conserve ribosomes for the following rapid restoration of their functioning as soon as normal growth conditions are restored. The functioning of ribosome hibernation factors can lead to the formation of a dormant state in a bacterial cell. The dormant state is a metabolically inactive state characterized by growth arrest (Prossliner et al., 2018; Trösch, Willmund, 2019; Usachev et al., 2020).

The formation of persistence is associated with dormancy. Persisters are rare variants of regular cells that have multidrug tolerance and are one of the reasons for the recalcitrance of chronic infectious diseases (Lewis, 2010; Zhang, 2014; Balaban et al., 2019). In addition, persisters are capable of mutating and surviving during exposure to high concentrations of antibiotics and, therefore, can be considered as a “reservoir” for the emergence of resistant mutants (Zhang, 2014; Tkachenko, 2018). Although the molecular mechanisms of persistence are still poorly understood, recently a model for the formation of persisters as a result of ribosome hibernation factors’ activity has been suggested (Song, Wood, 2020). There are also data on the involvement of polyamines and indole in persistence (Tkachenko et al., 2014, 2017; Zarkan et al., 2020; Lang et al., 2021). It can be assumed that these signaling molecules are involved in the formation of persistence through modulation of the expression of genes encoding ribosome hibernation factors. This is due to an ability of polyamines to stimulate the transcription of the *rmf*, *raiA*, *sra*, *ettA*, *rsfS* genes encoding ribosome hibernation factors, whereas indole is able to increase the transcription of two of them, *raiA* and *rmf* (Khaova et al., 2022).

These factors have different mechanisms of action. RMF (alternative names – Res, RimF) together with the HPF factor (alternative name – YhbH) can form inactive 100S dimers of ribosomes. RaiA (alternative names – YfiA, pY, Urf1) is able to block the active centers of 70S ribosomes, whereas SRA (alternative names – RpsV, Protein D) inhibits the translation

Table 1. *E. coli* strains used in this work

<i>E. coli</i> strain	Relevant characteristics	Source or reference
BW25141	F-, $\Delta(\text{araD-araB})567$, $\Delta\text{lacZ4787}::\text{rrnB-3}$, $\Delta(\text{phoBphoR})580$, $\lambda\text{-galU95}$, $\Delta\text{luidA3}::\text{pir+}$, recA1 , $\text{endA9}(\text{del-ins})::\text{FRT}$, rph1 , $\Delta(\text{rhaD-rhaB})568$, hsdR514	Datsenko, Wanner, 2000
BW25141 <i>rmf::lacZ</i>	As BW25141 but $\lambda\text{RS45 rmf39}::\text{lacZ}(\text{Hyb})$	This study
BW25141 <i>raiA::lacZ</i>	As BW25141 but $\lambda\text{RS45 raiA174}::\text{lacZ}(\text{Hyb})$	
BW25141 <i>sra::lacZ</i>	As BW25141 but $\lambda\text{RS45 sra102}::\text{lacZ}(\text{Hyb})$	
BW25141 <i>ettA::lacZ</i>	As BW25141 but $\lambda\text{RS45 ettA663}::\text{lacZ}(\text{Hyb})$	
BW25141 <i>rsfS::lacZ</i>	As BW25141 but $\lambda\text{RS45 rsfS81}::\text{lacZ}(\text{Hyb})$	

by interacting with the 30S subunit, and EttA (alternative name – YjjK) inactivates ribosomes in response to low intracellular levels of ATP. Finally, RsfS (alternative names – YbeB, RsfA) prevents the interaction of 50S and 30S subunits (Prossliner et al., 2018).

The aim of this work is to study the effects of indole on the expression of the *raiA* and *rmf* genes at the translational level, as well as the regulatory effects of the polyamines putrescine, cadaverine and spermidine on the translation of the *rmf*, *raiA*, *sra*, *ettA* and *rsfS* genes.

Materials and methods

Strains and growth conditions. *E. coli* strains used in this work are listed in Table 1. The cells of strains were grown in LB broth (Sigma) or defined medium M9 (+0.4 % glucose) in thermoshaker GFL-1092 (GFL) at 37 °C and 120 rpm. Media were supplemented with kanamycin 25 µg/ml (AppliChem) and/or ampicillin 50 µg/ml (AppliChem) when required. LB broth was used for strain constructions and routine cell growth.

For experiments studying the gene expression, cells of strains harboring *lacZ*-fusions were grown in defined medium M9 (+0.4 % glucose). Putrescine, cadaverine, and spermidine hydrochlorides (Sigma) were added to the medium for 2 h of cultivation at the concentrations indicated in the figures. Tryptophan (AppliChem) at 2 mM was added as previously described (Khaova et al., 2022).

Construction of the translational *lacZ*-fusions.

BW252141 strains with the chromosomal *lacZ*-fusions were obtained by using the pRS552/ λ RS45 system (Simons et al., 1987). Primers used in this work are listed in Table 2.

The in-between and resulting genetic constructs were verified by PCR and sequenced. Sequencing was performed by Evrogen (Moscow, Russia). All of the enzymes used were purchased from Thermo Fisher Scientific.

β -galactosidase assay. Gene expression was detected by the β -galactosidase activity by Miller's method (Miller, 1972).

mRNA secondary structure prediction. mRNA secondary structures were predicted using the RNAfold program (Lorenz et al., 2011).

Statistical analysis. Statistica for Windows 5.0 (StatSoft, Inc., 1995) software was used in the processing of experi-

Table 2. Primers used for construction of the translational *lacZ*-fusions

Gene	Sequence
<i>rmf</i>	5'-CGAATTCGGTATGTTGCCCTGAG-3' 5'-GTGGATCCTGTGCCCGTTC-3'
<i>raiA</i>	5'-ATAATCGGATCCCGTTTGGTCCGTATT-3' 5'-ACCAGAGGATCCTTAGGTGATTGAT-3'
<i>sra</i>	5'-ATGGATTGGAATCTTGCTCT-3' 5'-GGTTGGATCCTTTACTACGCT-3'
<i>ettA</i>	5'-TTTTGAATTCTACTGCGAGGGTGAT-3' 5'-CGGGATCCAGGAAGTAACGGT-3'
<i>rsfS</i>	5'-ATATTGAATTCGTCAGCCATCAGGGTGTA-3' 5'-TTGGGATCCACGCTAAGGCGATG-3'

mental data, presented as mean values of 3–5 independent experiments \pm standard deviation (Mean \pm SD).

Results

Polyamines are known to be able to regulate the gene expression. Genes, the expression of which is upregulated by polyamines at the translational level, are structured into the polyamine modulon. There are several mechanisms by which polyamines are able to modulate gene expression at the translational level, and, accordingly, specific features in the mRNA structure of such genes. Currently, there are data that mostly indicate the involvement of polyamines in the regulation of gene expression at the level of translational initiation. Firstly, these metabolites contribute to the initiation of the translation of genes, the mRNA structure of which contains the minor (ineffective) start codon. Secondly, polyamines stimulate the initiation of translation of genes, the mRNA structure of which has an unusually long distance between the Shine–Dalgarno sequence and the start codon. By introducing a bend in the mRNA at this region, polyamines are able to reduce this distance (Igarashi, Kashiwagi, 2006, 2018). Thirdly, polyamines are able to relax secondary structures such as “bulged-out” regions, which, being located between the Shine–Dalgarno sequence and the start codon, prevent the initiation of translation (Lightfoot, Hall, 2014).

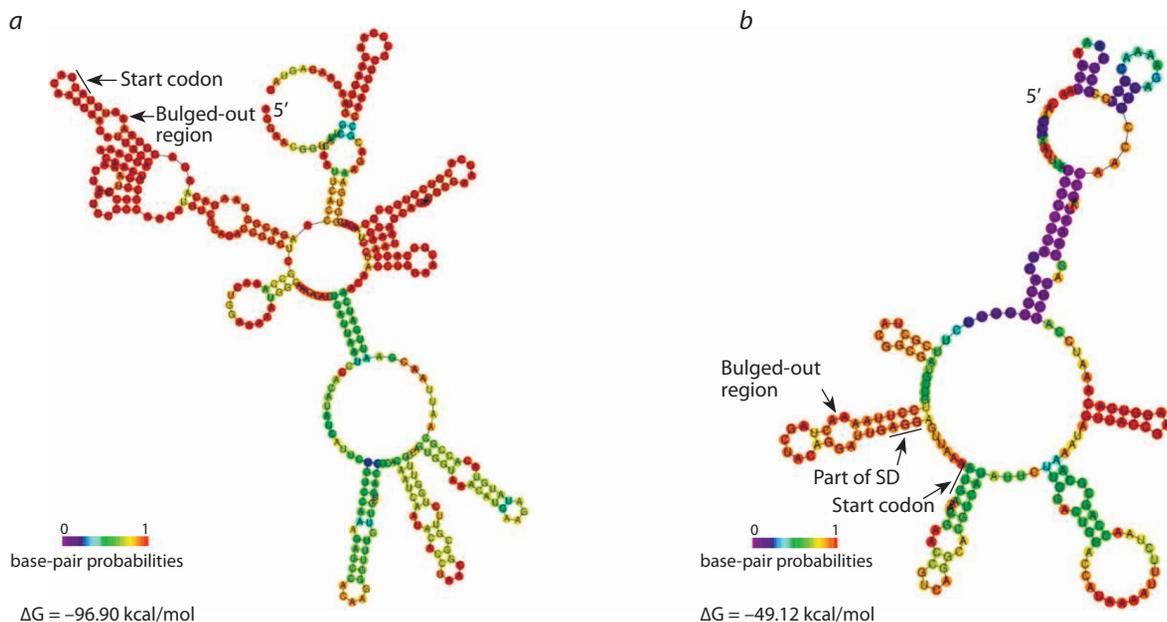


Fig. 1. Models of mRNA secondary structures of the *raiA* (a) and *sra* (b) genes obtained using the RNAfold program (Lorenz et al., 2011). SD – Shine–Dalgarno sequence.

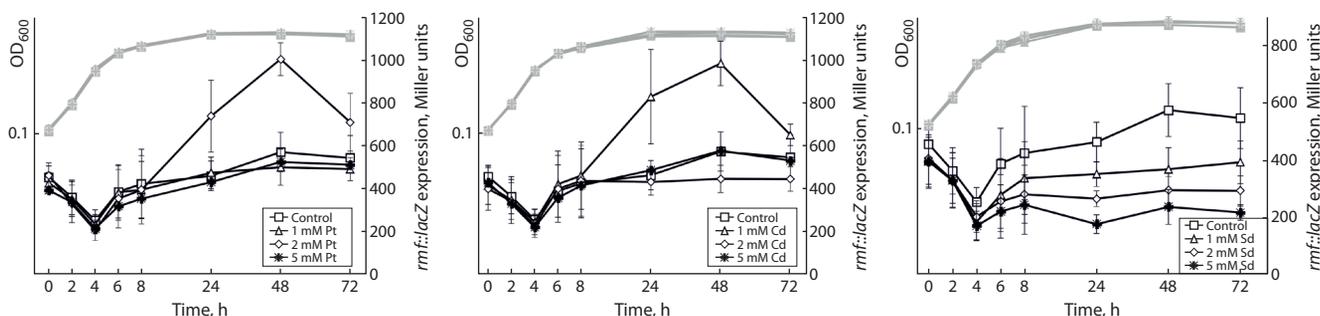


Fig. 2. The influence of polyamine additions – putrescine (Pt), cadaverine (Cd) and spermidine (Sd) – at different concentrations on the expression of the *rmf* gene at the translational level.

Here and elsewhere: gray curves – optical density, black curves – gene expression. OD₆₀₀ – optical density at 600 nm.

We have analyzed the mRNA structures of the *rmf*, *raiA*, *sra*, *ettA*, *rsfS* genes for the presence of the specific features of polyamine modulon genes using the GenBank and EcoCyc databases (Keseler et al., 2021) and published data, as well as the obtained models of mRNA secondary structures using the RNAfold program (Lorenz et al., 2011). According to published data, the *rmf* mRNA is characterized by an unusually long distance between the Shine–Dalgarno sequence and the start codon and the presence of a bulged-out structure in this region (Sakamoto et al., 2020). However, there is no information in the literature on the effect of polyamines on the expression of the remaining genes. According to the mRNA sequence of the genes, *ettA* mRNA and *rsfS* mRNA contain minor start codons GUG and UUG, respectively. The obtained models of mRNA secondary structures showed that for two genes, *raiA* and *sra*, the bulged-out structure can occur in the region of interest with a high probability (Fig. 1). The model of the *raiA* mRNA secondary structure demonstrates the presence

of a bulged-out structure at a distance of 3 nucleotides from the start codon. According to the model obtained for the *sra* mRNA secondary structure, the bulged-out region comprises a part of the Shine–Dalgarno sequence. Thus, the studied genes have properties specific for polyamine modulon.

Using obtained strains harboring *lacZ*-fusions, we studied the effects of polyamine additions at different concentrations – putrescine, cadaverine, and spermidine – on the expression of the *rmf*, *raiA*, *sra*, *ettA*, and *rsfS* genes at the translational level. The obtained results demonstrate that *rmf* gene expression is significantly stimulated by the addition of putrescine at 2 mM and cadaverine at 1 mM (Fig. 2). The maximal stimulatory effect is observed at 48 h of cultivation, when the bacterial cells are at the stationary phase. In contrast, spermidine inhibits *rmf* expression in proportion to the concentration of the supplement. Native *rmf* gene expression without additions remains at a consistently high level during the stationary phase.

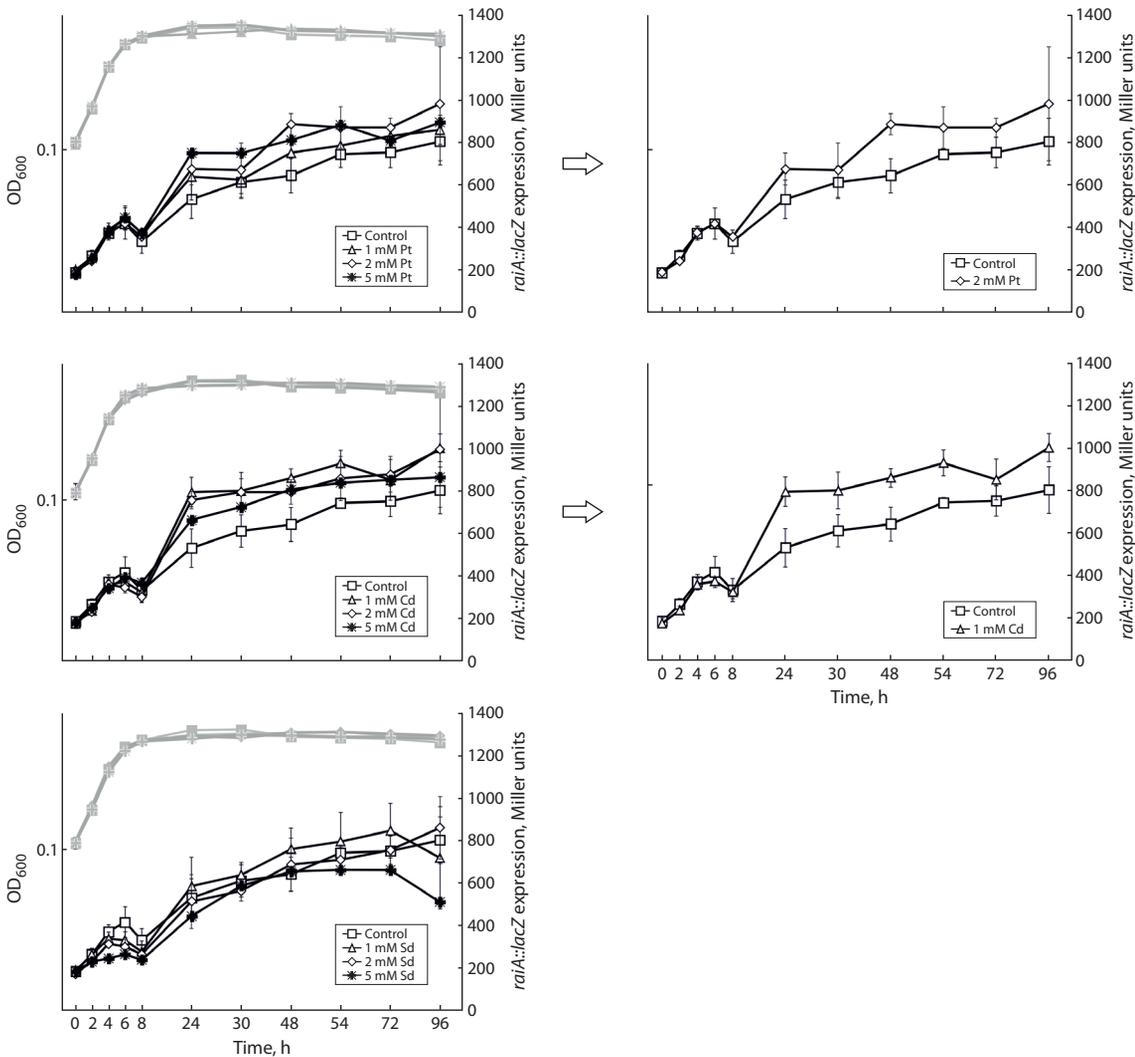


Fig. 3. The influence of polyamine additions – putrescine (Pt), cadaverine (Cd) and spermidine (Sd) – at different concentrations on the expression of the *raiA* gene at the translational level.

OD₆₀₀ – optical density at 600 nm.

According to the results obtained, the expression of the *raiA* gene is at a low level during exponential growth and increases at the stationary phase (Fig. 3). Supplements of putrescine at 2 mM and cadaverine at 1 mM insignificantly increase the expression of *raiA* in the stationary phase, whereas spermidine has no effect.

According to the obtained data, the expression of the *sra* gene is also at a constantly high level at the stationary phase (Fig. 4). The expression of *sra* is noticeably increased by the addition of cadaverine at 1 mM and 2 mM. In this case, the maximal effect is observed at the stationary phase (48 h of cultivation). Additions of putrescine and spermidine have no effect on the *sra* expression.

In contrast to the above-mentioned genes, the maximal expression of the *ettA* and *rsfS* genes is observed at the exponential phase (Fig. 5). The highest expression for *ettA* occurs at 4 h of cultivation, and for *rsfS*, at 1–3 h. Polyamine supplements have no effect on the expression of these genes.

We have previously shown that the addition of 2 mM tryptophan at 0 h of cultivation is equivalently converted into indole at 24 h. In this case, the indole content during 7 h was detected to be at a low level, similar to the control, and increased dramatically at 24 h (Khaova et al., 2022), because the gene of the tryptophanase TnaA, catalyzing the formation of indole from tryptophan, is expressed in a RpoS-dependent manner (Li, Young, 2013; Gaimster et al., 2014). Under these conditions, we have previously studied the expression of a number of genes responsible for adaptation of stress in *E. coli* at the transcriptional level. The expression of only two genes, *raiA* and *rmf*, was elevated in response to an increase in indole content (Khaova et al., 2022). In this regard, we investigated the effect of indole on the expression of these genes at the translational level under the same conditions (Fig. 6). The results demonstrate that despite the stimulatory effect at the transcriptional level, the expression of both of these genes at the translational level dropped with the increase in the indole content, starting from 24 h of observation.

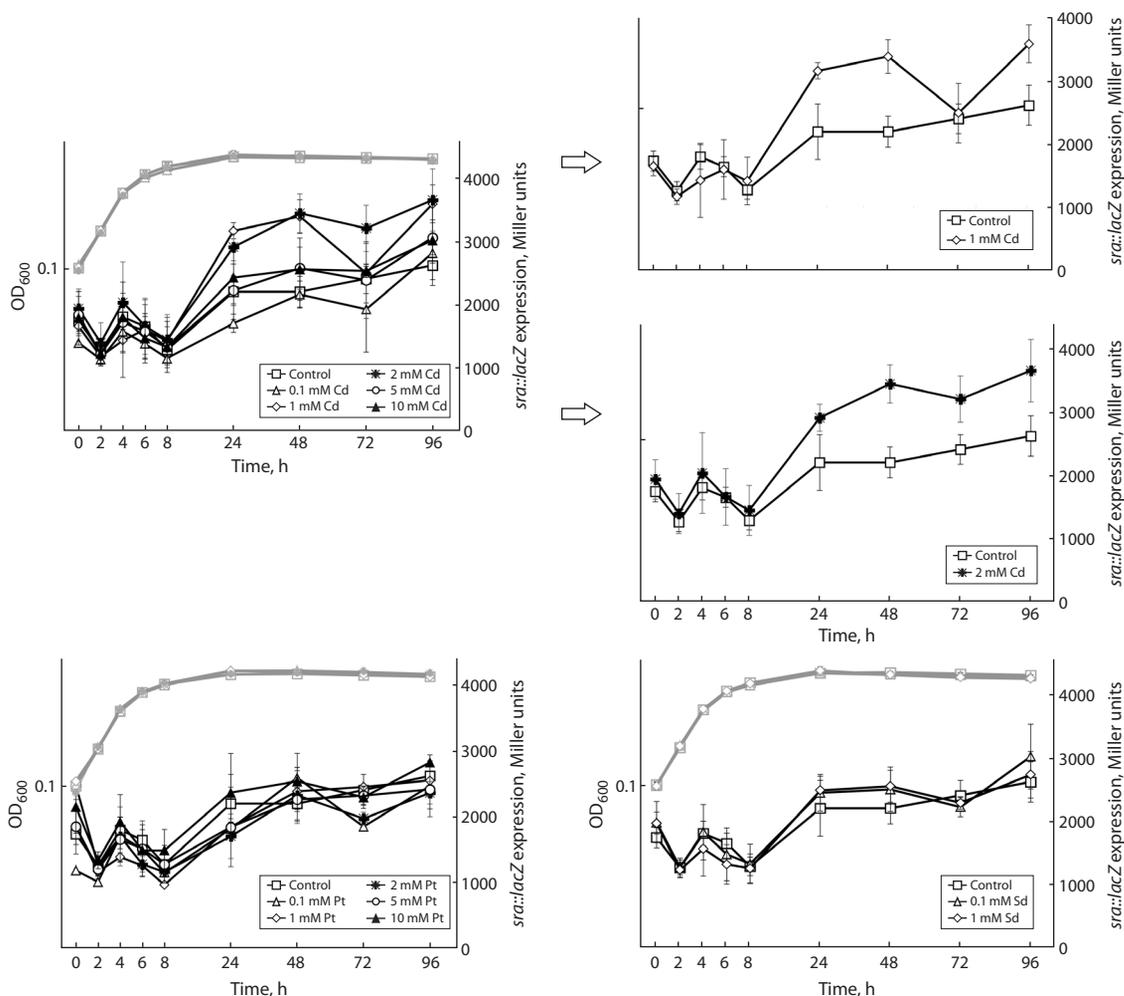


Fig. 4. The influence of polyamine additions – putrescine (Pt), cadaverine (Cd) and spermidine (Sd) – at different concentrations on the expression of the *sra* gene at the translational level.

OD₆₀₀ – optical density at 600 nm.

Discussion

The functions of ribosome hibernation factors are to reversibly inhibit such a resource intensive process as translation under starvation and other stresses. These factors are known to be under the control of master regulators (p)ppGpp, RpoS, CRP-cAMP responsible for the adaptation of bacteria to multiple stressors at the stationary phase. Due to this, ribosome hibernation factors predominantly function at this period. However, during exponential growth, these factors are also able to maintain a base level of inactive ribosomes (Prossliner et al., 2018). The results we obtained show that most expression of the *rmf*, *raiA* and *sra* genes is observed exactly at the stationary phase (Table 3). In contrast, the *ettA* and *rsfS* genes demonstrate the maximal expression during exponential growth. The obtained results show that polyamines affect the expression of the “stationary-phase genes” *rmf*, *raiA* and *sra*, but not *ettA* and *rsfS*. In all cases, the stimulatory effect of polyamines is observed at the stationary phase. Thus, polyamines are able to induce the expression of genes responsible for the adaptation to stress, and thereby contribute to the formation of an adaptive

state of a bacterial cell for the stationary phase. Moreover, the stimulatory effect is specific for the type of polyamine. The expression of each gene we studied depends on certain polyamines. The gene expression was predominantly positively modulated by putrescine and cadaverine.

Although indole stimulated the expression of the *rmf* and *raiA* genes at the transcriptional level (Khaova et al., 2022), the results obtained in this work show its significant inhibitory effect at the translational level under the same conditions. This may indicate the possibility of a post-transcriptional mechanism for regulation of gene expression.

Conclusion

The analysis of the mRNA primary structure of the *rmf*, *raiA*, *sra*, *ettA*, *rsfS* genes, encoding ribosome hibernation factors, as well as the obtained models of mRNA secondary structures showed that the studied genes have features of the polyamine modulon. We constructed strains harboring the translational *lacZ*-fusions and studied gene expression upon addition of the polyamines putrescine, cadaverine, and spermidine at differ-

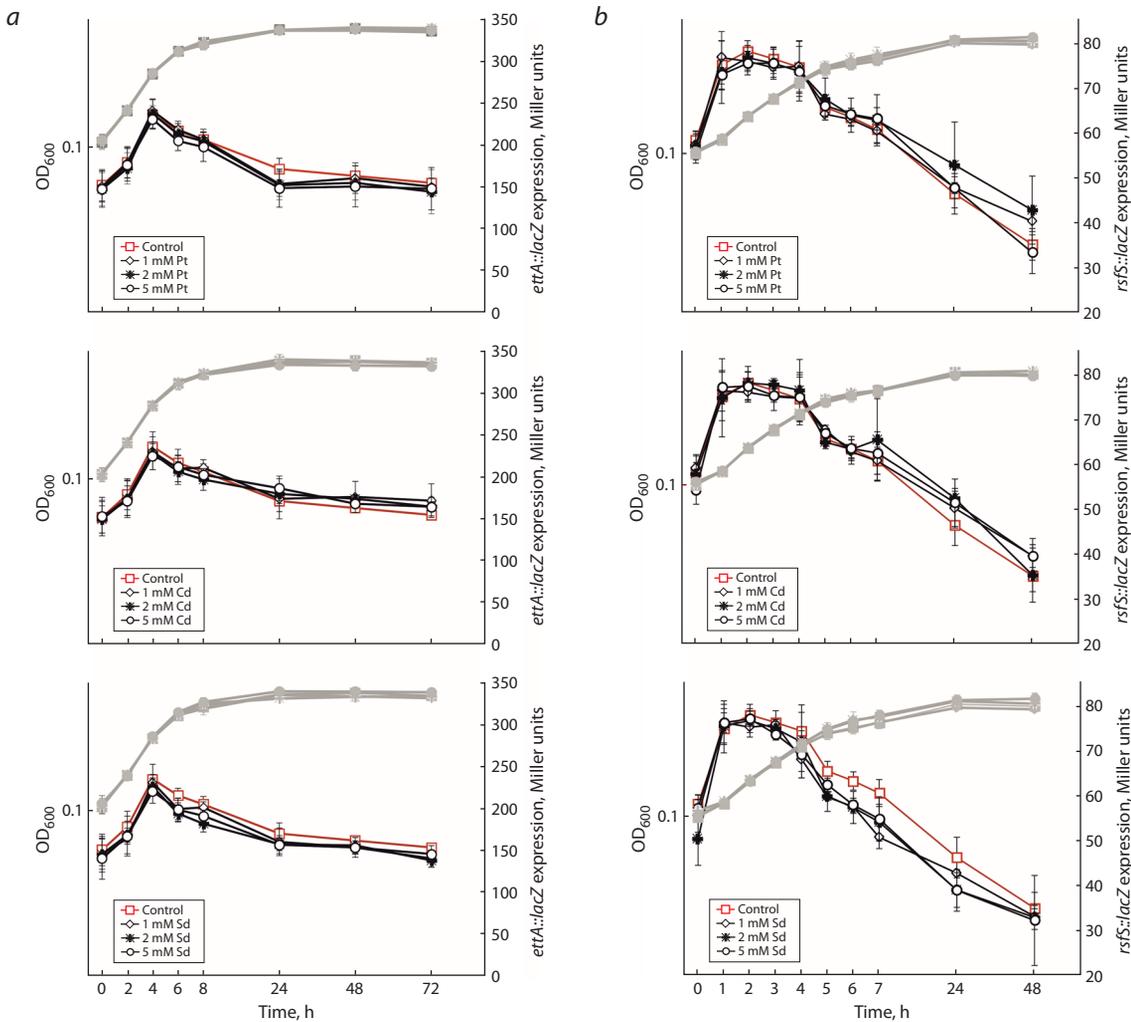


Fig. 5. The influence of polyamine additions – putrescine (Pt), cadaverine (Cd) and spermidine (Sd) – at different concentrations on the expression of the *ettA* (a) and *rsfS* (b) genes at the translational level. OD_{600} – optical density at 600 nm.

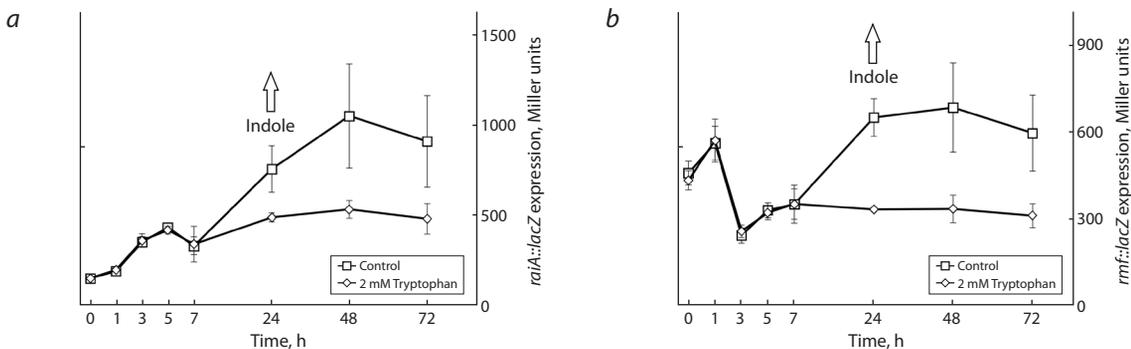


Fig. 6. The influence of indole on the expression of the *raiA* (a) and *rmf* (b) genes at the translational level.

ent concentrations. These genes, with the exception of *rmf*, were studied for the effect of polyamines on their expression for the first time. The stimulatory effect of polyamines was observed at the stationary phase and was specific to the type of polyamine. Polyamines affected the expression of the *rmf*,

raiA and *sra* genes, active at the stationary phase, but not *ettA* and *rsfS*, in which the highest expression was observed during exponential growth. Moreover, it was found that indole inhibits the expression of the *raiA* and *rmf* genes at the translational level, despite positive modulation at the transcriptional

Table 3. Effects of polyamines on the expression of the *rmf*, *raiA*, *sra*, *ettA*, *rsfS* genes at the translational level

Parameter	<i>rmf</i>			<i>raiA</i>			<i>sra</i>			<i>ettA</i>			<i>rsfS</i>		
	Pt	Cd	Sd	Pt	Cd	Sd	Pt	Cd	Sd	Pt	Cd	Sd	Pt	Cd	Sd
Effect of PA	+	+	-	+	+	No	No	+	No	No	No	No	No	No	No
Period of maximal expression	Stationary phase			Stationary phase			Stationary phase			Exponential phase			Exponential phase		
Features of PA modulon	Bulged-out region, unusually long distance between SD and start codon			Bulged-out region			Bulged-out region			Minor start codon			Minor start codon		

Note. "+" – stimulating effect, "-" – inhibitory effect, "no" – no effect. PA – polyamine, Pt – putrescine, Cd – cadaverine, Sd – spermidine, SD – Shine-Dalgarno sequence.

level, which may indicate the possibility of a post-transcriptional regulation of the gene expression. The results obtained open up prospects for further research in this direction.

References

- Balaban N.Q., Helaine S., Lewis K., Ackermann M., Aldridge B., Andersson D.I., Brynildsen M.P., Bumann D., Camilli A., Collins J.J., Dehio C., Fortune S., Ghigo J.-M., Hardt W.-D., Harms A., Heineemann M., Hung D.T., Jenal U., Levin B.R., Michiels J., Storz G., Tan M.-W., Tenson T., Melderen L.V., Zinkernagel A. Definitions and guidelines for research on antibiotic persistence. *Nat. Rev. Microbiol.* 2019;17(7):441-448. DOI 10.1038/s41579-019-0196-3
- Datsenko K.A., Wanner B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA.* 2000;97(12):6640-6645. DOI 10.1073/pnas.120163297
- Gaimster H., Cama J., Hernández-Ainsa S., Keyser U.F., Summers D.K. The indole pulse: a new perspective on indole signalling in *Escherichia coli*. *PLoS One.* 2014;9(4):e93168. DOI 10.1371/journal.pone.0093168
- Igarashi K., Kashiwagi K. Polyamine modulon in *Escherichia coli*: genes involved in the stimulation of cell growth by polyamines. *J. Biochem.* 2006;139(1):11-16. DOI 10.1093/jb/mvj020
- Igarashi K., Kashiwagi K. Effects of polyamines on protein synthesis and growth of *Escherichia coli*. *J. Biol. Chem.* 2018;293(48):18702-18709. DOI 10.1074/jbc.TM118.003465
- Keseler I.M., Gama-Castro S., Mackie A., Billington R., Bonavides-Martinez C., Caspi R., Kothari A., Krummenacker M., Midford P.E., Muñoz-Rascado L., Ong W.K., Paley S., Santos-Zavaleta A., Subhraveti P., Tierrafria V.H., Wolfe A.J., Collado-Vides J., Paulsen I.T., Karp P.D. The EcoCyc database in 2021. *Front. Microbiol.* 2021;12:711077. DOI 10.3389/fmicb.2021.711077
- Khaova E.A., Kashevarova N.M., Tkachenko A.G. Regulatory effect of polyamines and indole on expression of stress adaptation genes in *Escherichia coli*. *Acta Biomedica Scientifica.* 2022;7(3):150-161. DOI 10.29413/ABS.2022-7.3.16 (in Russian)
- Kim C.S., Li J.H., Barco B., Park H.B., Gatsios A., Damania A., Wang R., Wyche T.P., Piizzi G., Clay N.K., Crawford J.M. Cellular stress upregulates indole signaling metabolites in *Escherichia coli*. *Cell Chem. Biol.* 2020;27(6):698-707.e7. DOI 10.1016/j.chembiol.2020.03.003
- Kusano T., Berberich T., Tateda C., Takahashi Y. Polyamines: essential factors for growth and survival. *Planta.* 2008;228(3):367-381. DOI 10.1007/s00425-008-0772-7
- Lang M., Krin E., Korlowski C., Sismeiro O., Varet H., Coppée J.Y., Mazel D., Baharoglu Z. Sleeping ribosomes: bacterial signaling triggers RaiA mediated persistence to aminoglycosides. *iScience.* 2021;24(10):103128. DOI 10.1016/j.isci.2021.103128
- Lee J.H., Wood T.K., Lee J. Roles of indole as an interspecies and interkingdom signaling molecule. *Trends Microbiol.* 2015;23(11):707-718. DOI 10.1016/j.tim.2015.08.001
- Lewis K. Persister cells. *Annu. Rev. Microbiol.* 2010;64:357-372. DOI 10.1146/annurev.micro.112408.134306
- Li G., Young K.D. Indole production by the tryptophanase TnaA in *Escherichia coli* is determined by the amount of exogenous tryptophan. *Microbiology (Reading).* 2013;159(2):402-410. DOI 10.1099/mic.0.064139-0
- Lightfoot H.L., Hall J. Endogenous polyamine function – the RNA perspective. *Nucleic Acids Res.* 2014;42(18):11275-11290. DOI 10.1093/nar/gku837
- Lorenz R., Bernhart S.H., Tafer H., Flamm C., Stadler P.F., Hofacker I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 2011;6:26. DOI 10.1186/1748-7188-6-2
- Michael A.J. Polyamines in eukaryotes, bacteria, and archaea. *J. Biol. Chem.* 2016;291(29):14896-14903. DOI 10.1074/jbc.R116.734780
- Michael A.J. Polyamine function in archaea and bacteria. *J. Biol. Chem.* 2018;293(48):18693-18701. DOI 10.1074/jbc.TM118.005670
- Miller J.H. Experiments in Molecular Genetics. New York, 1972
- Miller-Fleming L., Olin-Sandoval V., Campbell K., Ralser M. Remaining mysteries of molecular biology: the role of polyamines in the cell. *J. Mol. Biol.* 2015;427(21):3389-3406. DOI 10.1016/j.jmb.2015.06.020
- Nesterova L.Y., Akhova A.V., Tkachenko A.G. Influence of indole on intracellular polyamines and antibiotic susceptibility of *Escherichia coli*. *Vestnik Moskovskogo universiteta. Seriya 16. Biologiya = Herald of Moscow University. Series 16. Biology.* 2021;76(4):219-224 (in Russian)
- Prossliner T., Skovbo Winther K., Sørensen M.A., Gerdes K. Ribosome hibernation. *Annu. Rev. Genet.* 2018;52:321-348. DOI 10.1146/annurev-genet-120215-035130
- Rhee H.J., Kim E.J., Lee J.K. Physiological polyamines: simple primordial stress molecules. *J. Cell. Mol. Med.* 2007;11(4):685-703. DOI 10.1111/j.1582-4934.2007.00077.x
- Sakamoto A., Sahara J., Kawai G., Yamamoto K., Ishihama A., Uemura T., Igarashi K., Kashiwagi K., Terui Y. Cytotoxic mechanism of excess polyamines functions through translational repression of specific proteins encoded by polyamine modulon. *Int. J. Mol. Sci.* 2020;21(7):2406. DOI 10.3390/ijms21072406
- Shah P., Swiatlo E. A multifaceted role for polyamines in bacterial pathogens. *Mol. Microbiol.* 2008;68(1):4-16. DOI 10.1111/j.1365-2958.2008.06126.x
- Simons R.W., Houman F., Kleckner N. Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene.* 1987;53(1):85-96. DOI 10.1016/0378-1119(87)90095-3
- Song S., Wood T.K. ppGpp ribosome dimerization model for bacterial persister formation and resuscitation. *Biochem. Biophys. Res. Commun.* 2020;523(2):281-286. DOI 10.1016/j.bbrc.2020.01.102

- Tabor C.W., Tabor H. Polyamines in microorganisms. *Microbiol. Rev.* 1985;49(1):81-99. DOI 10.1128/mr.49.1.81-99.1985
- Tkachenko A.G. Molecular Mechanisms of Stress Responses in Microorganisms. Yekaterinburg, 2012 (in Russian)
- Tkachenko A.G. Stress responses of bacterial cells as mechanisms of development of antibiotic tolerance (review). *Applied Biochemistry and Microbiology.* 2018;54(2):108-127. DOI 10.1134/S0003683818020114
- Tkachenko A.G., Akhova A.V., Shumkov M.S., Nesterova L.Yu. Polyamines reduce oxidative stress in *Escherichia coli* cells exposed to bactericidal antibiotics. *Res. Microbiol.* 2012;163(2):83-91. DOI 10.1016/j.resmic.2011.10.009
- Tkachenko A.G., Kashevarova N.M., Karavaeva E.A., Shumkov M.S. Putrescine controls the formation of *Escherichia coli* persister cells tolerant to aminoglycoside netilmicin. *FEMS Microbiol. Lett.* 2014;361(1):25-33. DOI 10.1111/1574-6968.12613
- Tkachenko A.G., Kashevarova N.M., Tyuleneva E.A., Shumkov M.S. Stationary-phase genes upregulated by polyamines are responsible for the formation of *Escherichia coli* persister cells tolerant to netilmicin. *FEMS Microbiol. Lett.* 2017;364(9):fmx084. DOI 10.1093/femsle/fmx084
- Trösch R., Willmund F. The conserved theme of ribosome hibernation: from bacteria to chloroplasts of plants. *Biol. Chem.* 2019;400(7):879-893. DOI 10.1515/hsz-2018-0436
- Usachev K.S., Yusupov M.M., Validov Sh.Z. Hibernation as a stage of ribosome functioning. *Biochemistry (Moscow).* 2020;85(11):1434-1442. DOI 10.1134/S0006297920110115
- Zarkan A., Liu J., Matuszewska M., Gaimster H., Summers D.K. Local and universal action: the paradoxes of indole signalling in bacteria. *Trends Microbiol.* 2020;28(7):566-577. DOI 10.1016/j.tim.2020.02.007
- Zhang Y. Persisters, persistent infections and the Yin-Yang model. *Emerg. Microbes Infect.* 2014;3(1):e3. DOI 10.1038/emi.2014.3

ORCID

E.A. Khaova orcid.org/0000-0003-4457-2652
A.G. Tkachenko orcid.org/0000-0002-8631-8583

Acknowledgements. The work was financially supported by the Ministry of Science and Higher Education of the Russian Federation (No. 124020500028-4).

Conflict of interest. The authors declare no conflict of interest.

Received July 14, 2023. Revised November 13, 2023. Accepted November 15, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Expression of auxin transporter genes in flax (*Linum usitatissimum*) fibers during gravity response

N.N. Ibragimova , N.E. MokshinaKazan Institute of Biochemistry and Biophysics of Kazan Scientific Center of the Russian Academy of Sciences, Kazan, Russia
 nibra@yandex.ru

Abstract. Gravitropism is an adaptive reaction of plants associated with the ability of various plant organs to be located and to grow in a certain direction relative to the gravity vector, while usually the asymmetric distribution of the phytohormone auxin is a necessary condition for the gravitropical bending of plant organs. Earlier, we described significant morphological changes in phloem fibers with a thickened cell wall located on different sides of the stem in the area of the gravitropic curvature. The present study is the first work devoted to the identification of genes encoding auxin transporters in cells at different stages of development and during gravity response. In this study, the flax genes encoding the AUX1/LAX, PIN-FORMED, PIN-LIKES, and ABCB auxin transporters were identified. A comparative analysis of the expression of these genes in flax phloem fibers at different stages of development revealed increased expression of some of these genes at the stage of intrusive growth (*LusLAX2* (A, B), *LuxPIN1-D*, *LusPILS7* (C, D)), at the early stage of tertiary cell wall formation (*LusAUX1* (A, D), *LusABCB1* (A, B), *LusABCB15-A*, *LusPIN1* (A, B), *LusPIN4-A*, and *LusPIN5-A*), and at the late stage of tertiary cell wall development (*LusLAX3* (A, B)). It was shown that in the course of gravitropism, the expression of many genes, including those responsible for the influx of auxin in cells (*LusAUX1-D*), in the studied families increased. Differential expression of auxin transporter genes was revealed during gravity response in fibers located on different sides of the stem (upper (PUL) and lower (OPP)). The difference was observed due to the expression of genes, the products of which are responsible for auxin intracellular transport (*LusPILS3*, *LusPILS7-A*) and its efflux (*LusABCB15-B*, *LusABCB19-B*). It was noted that the increased expression of *PIN* genes and *ABCB* genes was more typical of fibers on the opposite side. The results obtained allow us to make an assumption about the presence of differential auxin content in the fibers of different sides of gravistimulated flax plants, which may be determined by an uneven outflow of auxin. This study gives an idea of auxin carriers in flax and lays the foundation for further studies of their functions in the development of phloem fiber and in gravity response.

Key words: flax; *Linum usitatissimum* L.; gravitropism; fiber; auxin transport; gene expression.

For citation: Ibragimova N.N., Mokshina N.E. Expression of auxin transporter genes in flax (*Linum usitatissimum*) fibers during gravity response. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(1): 33-43. DOI 10.18699/VJGB-24-05

Экспрессия генов транспортеров ауксина в волокнах льна (*Linum usitatissimum*) при гравитответе

Н.Н. Ибрагимова , Н.Е. МокшинаКазанский институт биохимии и биофизики – обособленное структурное подразделение Федерального исследовательского центра «Казанский научный центр Российской академии наук», Казань, Россия
 nibra@yandex.ru

Аннотация. Гравитропизм – адаптивная реакция растений, связанная со способностью органов растений располагаться и расти в определенном направлении относительно вектора силы тяжести. При этом асимметричное распределение фитогормона ауксина считается необходимым условием для тропического изгиба органов растения. Ранее нами были описаны яркие морфологические изменения флоэмных волокон с утолщенной клеточной стенкой, находящихся на разных сторонах зрелых участков стебля в области гравитропического изгиба. Настоящее исследование – первая работа, посвященная идентификации генов, кодирующих переносчики ауксина в этих клетках на разных стадиях развития и при гравитответе. В растениях льна идентифицированы гены основных переносчиков ауксина: AUX1/LAX, PIN-FORMED, PIN-LIKES и ABCB. Сравнительный анализ экспрессии этих генов во флоэмных волокнах льна, находящихся на разных стадиях развития, выявил повышенную экспрессию некоторых генов на стадии интрузивного роста (*LusLAX2* (A, B), *LuxPIN1-D*, *LusPILS7* (C, D)), на ранней стадии формирования третичной клеточной стенки (*LusAUX1* (A, D), *LusABCB1* (A, B), *LusABCB15-A*, *LusPIN1* (A, B), *LusPIN4-A*, *LusPIN5-A*) и на поздней стадии развития третичной клеточной стенки (*LusLAX3* (A, B)). Показано, что при гравитропизме повышалась экспрессия многих генов исследуемых семейств, в том числе отвечающих за приток ауксина в клетки (*LusAUX1-D*). Выявлена дифференциальная экспрессия генов переносчиков ауксина

при гравитовете в волокнах, находящихся на разных сторонах стебля – верхней (PUL) и нижней (OPP): различие наблюдалось за счет экспрессии генов, продукты которых отвечают за внутриклеточный транспорт (*LusPILS3*, *LusPILS7-A*) и отток ауксина из клеток (*LusABCB15-B*, *LusABCB19-B*). Повышенная экспрессия *PIN*-генов и *ABCB*-генов была более типична для волокон OPP-стороны стебля. Полученные результаты позволяют сделать предположение о наличии дифференциального содержания ауксина в волокнах разных сторон стебля гравитостимулированных растений льна, которое, возможно, определяется неравномерным оттоком ауксина. Исследование дает представление о переносчиках ауксина во льне и закладывает основу для дальнейшего изучения их функций в развитии флоэмного волокна и при гравитовете.

Ключевые слова: лен; *Linum usitatissimum* L.; гравитропизм; волокно; транспорт ауксина; экспрессия генов.

Introduction

Plants are under the constant influence of abiotic and biotic factors, including unfavorable ones. The activity and coordinated action of auxin (indole-3-acetic acid, IAA) carriers in plants underlie a flexible network that mobilizes IAA in response to various environmental changes. This applies to plant tropisms as well. Regardless of the mechanisms involved, the activity of the phytohormone auxin is crucial for all tropisms, including gravitropism (Harrison, Pickard, 1989; Evans, 1991; Li et al., 1991; Rakusová et al., 2019). It has been shown that inhibitors of IAA transport block the development of the gravitropic reaction in plants (Li et al., 1991). The uneven distribution of auxin due to the participation of protein carriers during gravitropism is an important area of both fundamental and applied research related to plant lodging.

The implementation of the gravity response is associated with the formation of a gravitropic curvature. The formation of such curvature in plants occurs with the participation of different mechanisms: in young, actively growing organs (seedling roots, hypocotyles, coleoptiles), the bend is formed due to different rates of cell elongation on the different sides of the gravistimulated organ (Harrison, Pickard, 1989; Li et al., 1991; Zhu et al., 2019). On the other hand, the gravitropic curvature of mature stems, in addition to the above-mentioned mechanism (which, apparently, continues in the plant growing tip) in the parts of the stem that have ceased elongation, occurs due to the probable “contractile” properties of the fibers (Ibragimova et al., 2017). The mechanism of formation of gravitropic curvature of mature stems also has its own characteristics in plant species of different systematic groups: in angiosperm woody plants, it is formed due to changes in cambial activity and the formation of a tertiary cell wall (TCW) in xylem fibers on the upper side of the gravistimulated organ, leading to the formation of tension wood (Haygreen, Bowyer, 1996; Jourez et al., 2001). In mature gymnosperm stems, gravitropic curvature is provided by the formation of compression wood, which appears on the underside of the gravistimulated organ (Timell, 1969). Finally, in mature annual stems of herbaceous plants, including flax, gravitropic curvature is provided, as we assume, by primary phloem fibers having a cell wall (or TCW), while the formation of TCW is also observed in the xylem fibers of the upper stem side (Ibragimova et al., 2017). If, in the case of young organs, the role of auxin in the formation of curvature is actively studied, information on the distribution of IAA in mature organs is very limited and contradictory (Hellgren et al., 2004; Gerttula et al., 2015).

Auxin is distributed in the plant body by two different but interconnected transport systems: first, rapid flow in the

phloem together with photosynthetic assimilates; and second, slow and directional polar transport of auxin from cell to cell (Adamowski, Friml, 2015). While phloem transport provides a general way of auxin delivery from its place of synthesis to the recipient organs, polar transport distributes auxin in an accurate manner, which is critical for the formation of local auxin maxima and is one of the key elements in its functioning (Friml et al., 2002; Zažimalová et al., 2010; Adamowski, Friml, 2015). Auxin carriers of the PIN family form the main part of this system, controlling the direction and speed of transport through a number of cells (Zažimalová et al., 2010). As for possible changes in the expression of genes for the PIN protein, the relevant data are currently limited to several model species.

In addition to the PIN (PIN-FORMED) family, auxin transport is carried out by other types of proteins: AUX1/LAX (AUXIN-INSENSITIVE1/LIKE AUX1), ABCB (subfamily of ATP-binding cassette transporters), PILS (PIN-LIKES), NRT1.1 (nitrate transporter 1.1), and WAT1 (WALLS ARE THIN1) (Manna et al., 2022). It is believed that at a low pH of the apoplast, auxin becomes protonated and can penetrate into the cell by diffusion. In certain types of cells, auxin can be transported to the cytosol by protein carriers, members of the AUX1/LAX family (Swarup, Péret, 2012). Inside the cell, auxin becomes negatively charged, and consequently, carriers are required to ensure its efflux, such as PIN and ABCB, through the cell membrane into the apoplast (Zažimalová et al., 2010). A less-characterized group of PILS transport proteins is probably responsible for intracellular auxin transport (Barbez et al., 2012).

In this study, the genes for the main auxin carrier proteins (PIN, AUX1/LAX, ABCB, and PILS) were identified in the flax genome. Their expression was evaluated using comparative transcriptomic analysis of the phloem fibers, which were sampled from control and gravistimulated flax plants. As a model system, we selected flax phloem fibers at different stages of development (with primary (PCW) and thickened tertiary cell walls (TCW)), as well as phloem fibers from different sides of gravistimulated flax plants at a late stage of TCW development. Flax phloem fibers are arranged along the stem axis in the bundles, which simplifies their isolation at different stages of development, separated in space and time (fibers reach a finite length during intrusive growth, and then layers of secondary and tertiary CW are sequentially deposited in cells (Gorshkova et al., 2003)). All this makes it possible to conduct diverse studies at the cell level. It was shown that during flax gravistimulation, phloem fibers localized on different sides of the stem (upper (PUL) and lower (OPP)) had morphological

and structural biochemical changes (Ibragimova et al., 2017, 2020). Analysis of the expression of genes encoding the main auxin carriers in isolated fibers will reveal the type of auxin transport, which, as we assume, is activated in phloem fibers during gravity response.

Materials and methods

Identification of auxin transporters. Using the Phytozome database, protein sequences containing functional domains (Pfam) PF01490, PF03547 and PF03547, PF00005, characteristic of AUX1/LAX, PIN/PILS, and ABCB auxin transporters, respectively, were identified. The identified genes for auxin carriers in flax were named in accordance with the orthologous sequence of *Arabidopsis thaliana* (thale cress); the functions of the products of all identified genes are predicted since their annotation is based on homology to the characterized genes of the thale cress. All sequences of the analyzed genes are presented in an Supplementary Material¹.

Gene expression level and phylogenetic analysis. To evaluate gene expression, we used previously obtained transcriptomic data for flax plants (rapid growth phase), which are available in the FIBexDB database (<https://ssl.cres-t.org/fibex/flax/>) (Mokshina et al., 2021). For analysis, phloem fibers were taken at different stages of development: before the formation of TCW (the stage of intrusive growth, iFIBa), at the early stage of TCW formation (tFIBa), at the late stage of TCW formation (tFIBb), as well as on different sides of the stem (PUL – upper part of bending plants, and OPP – opposite part) at the late stage of TCW formation during gravistimulation. Gravistimulation was carried out by tilting the plants (at the base) parallel to the soil (90 degrees). Gene expression in fibers at the late stage of TCW formation was analyzed 8, 24, and 96 hours after the plants were inclined. More than two-fold changes are being discussed.

To build a phylogenetic tree, the Maximum Likelihood method was used, the Le_Gascuel_2008 model (LG+G); Bootstrap support 1000. It was performed in the MEGA7 program.

Results

The genes encoding the main families of auxin transporters were identified: AUX1/LAX, responsible for the influx of auxin into the cell (Swarup, Péret, 2012); PIN-FORMED and ABCB, responsible mainly for the auxin outflow (Zažímalová et al., 2010); PIN-LIKES (PILS), responsible for intracellular auxin transport (Barbez et al., 2012); expression of the listed genes was analyzed.

Identification and expression of *LusPINs* and *LusPILS*

In the flax genome, when searching in the Phytozome database (<https://phytozome-next.jgi.doe.gov/>) according to the presence of the PF03547 membrane transporter domain, 34 genes were found that correspond to 12 orthologs in *Arabidopsis*; a total of 15 genes with PF03547 (8 – *PIN*, 7 – *PILS*) were found in the *Arabidopsis* genome.

The sequence of the *Lus10020829* gene (*AT2G01420*, *PIN4*) was corrected by us in the Augustus program (<https://bioinf.uni-greifswald.de/augustus/>) and continued by the

Lus10020830 sequence. The adjustment also took into account the results of BLASTX Arabi/Clami/Rice, available in the Phytozome database (JBrowse) (<https://phytozome-next.jgi.doe.gov/jbrowse/index.html>). Similarly, six more sequences were edited: *Lus10009685* (*AT1G73590*, *PIN1*), *Lus10002280* (*AT1G71090*, *PILS2*), *Lus10018006* (*AT1G73590*, *PIN1*), *Lus10036000* (*AT1G71090*, *PILS2*); *Lus100360001* was excluded from analysis), *Lus10042003* (*AT1G73590*, *PIN1*), and *Lus10016704* (*AT1G71090*, *PILS2*). The *Lus10004059* sequence (*AT1G71090*, *PILS2*) included a fragment of 119 unidentified amino acids (out of 443), which is probably due to problems with initial sequencing and/or genome assembly. Redundant domains were removed from the *Lus10016688* sequence, after which the closest homologue to it was established as *PILS7* (*AT5G65980*). Not typical domains were removed from the *Lus10019229* sequence (*AT1G20925*, *PILS1*), but it cannot be considered fully predicted since it failed to establish the position of the start codon *in silico*. The *Lus10012680* sequence (*AT2G01420*, *PIN4*) was increased from 231 to 517 amino acids. Sequences without predicted transmembrane domains were excluded from the analysis. When correcting the sequences, the integrity of some domains was restored, and the number of transmembrane domains and molecular weight approached the indicators characteristic of the members of the analyzed family.

Thus, after bioinformatics analysis and correction, 27 sequences remained out of 34, annotated as *PIN* membrane transporters, or *PILS*, which we used for further analysis (Table 1).

The molecular weight of these proteins varied from 33.5 to 75 kDa, the value of pI – from 5.3 to 9.6; the number of transmembrane domains – from 5 to 10 (see Table 1).

To annotate the *PIN/PILS* genes, in addition to the BLAST results, we performed a phylogenetic analysis of the amino acid sequences of *PIN/PILS* in *Arabidopsis* and flax (Fig. 1). The analyzed sequences were expected to be divided into two clades: *PIN* and *PILS*. Several orthologs of the *Arabidopsis* *PIN* corresponded to two paralog genes in the flax genome (*PIN2*, 5, 8); four flax sequences were in the same group with the *A. thaliana* *PIN1*, and two *L. usitatissimum* sequences corresponded to the *AtPIN3/4/7* group. Many *PILS* were also duplicated (see Fig. 1).

The expression of the genes *Lus10001429* (*LusPIN2-A*), *Lus10001637* (*LusPIN2-B*), *Lus10004287* (*LusPILS1-A*), *Lus10002280* (*LusPILS2-A*), and *Lus10010303* (*PIL8-A*) was low (<16 TGR) and was not further analyzed. At different stages of development and during gravistimulation, 22 *PIN/PILS* were expressed in the fibers. According to the dynamics of expression, several groups of genes can be distinguished. Interestingly, three genes (*LusPIN1-D*, *LusPILS7-C*, and *D*) showed an increased level of expression in fibers only at the stage of their elongation, while expression decreased in mature fibers, and there were no differences in fibers during gravistimulation (Fig. 2).

A group of genes was also revealed that had a pronounced expression peak in fibers at an early stage of TCW formation (tFIBa) (Fig. 3, *LusPIN1-A, B, LusPIN4-A*) or almost the same level in fibers during elongation and formation of thickened TCW (see Fig. 3, *LusPIN1-C, LusPIN4-B, LusPILS2-B*). At the same time, the expression of all these genes decreased in

¹ Supplementary Material is available at: <https://vavilovj-icg.ru/download/pict-2024-28/appx1.xlsx>

Table 1. List and some characteristics of LusPIN and LusPILS sequences

LUS	AT	LUS Name	AA	Mw, kDa	pI	TMH_LUS	TMH_AT
Lus10009054	AT1G73590	PIN1-A	613	66.0	8.9	9	9
Lus10009685/86	AT1G73590	PIN1-B	510	55.1	9.3	9	9
Lus10018006/07	AT1G73590	PIN1-C	498	53.7	9.0	9	9
Lus10042003/04	AT1G73590	PIN1-D	568	61.5	8.8	7	9
Lus10001429	AT5G57090	PIN2-A	650	69.5	9.1	9	9
Lus10001637	AT5G57090	PIN2-B	648	69.2	9.2	9	9
Lus10012680	AT2G01420	PIN4-A	517	57.0	8.8	7	10
Lus10020829/30	AT2G01420	PIN4-B	528	56.8	9.6	8	10
Lus10020193	AT5G16530	PIN5-A	354	38.8	7.4	9	9
Lus10026994	AT5G16530	PIN5-B	354	38.9	8.0	9	9
Lus10010303	AT5G15100	PIN8-A	359	39.3	9.3	8	8
Lus10013422	AT5G15100	PIN8-B	359	39.1	9.4	8	8
Lus10004287	AT1G20925	PILS1-A	417	44.5	8.3	9	9
Lus10016046	AT1G20925	PILS1-B	434	46.7	8.2	8	9
Lus10019229*	AT1G20925	PILS1-C	309	33.5	8.9	5	9
Lus10030715	AT1G20925	PILS1-D	394	42.8	9.3	8	9
Lus10002280/81	AT1G71090	PILS2-A	388	40.8	5.9	6	8
Lus10004059*	AT1G71090	PILS2-B	442	48.7	?	7	8
Lus10016704/05	AT1G71090	PILS2-C	465	51.1	5.3	9	8
Lus10036000/01	AT1G71090	PILS2-D	508	55.8	6.6	8	8
Lus10025166	AT1G76520	PILS3	406	43.5	7.8	9	10
Lus10003240	AT5G01990	PILS6-A	692	75.3	5.9	8	10
Lus10035610	AT5G01990	PILS6-B	422	45.3	8.2	9	10
Lus10001303	AT5G65980	PILS7-A	396	42.7	6.5	8	8
Lus10012708	AT5G65980	PILS7-B	396	42.7	6.3	8	8
Lus10016688	AT5G65980	PILS7-C	408	44.5	6.1	8	8
Lus10035978	AT5G65980	PILS7-D	392	42.8	6.7	8	8

Note. The corrected sequences are highlighted in bold. * Incomplete sequences.

Hereinafter: LUS – *Linum usitatissimum*; AT – *Arabidopsis thaliana*; AA – number of amino acids; Mw – molecular weight, kDa; pI – isoelectric point; TMH – number of transmembrane domains.

mature fibers and was activated again during gravistimulation, especially in OPP samples. The peak of expression during gravistimulation occurred at 24 hours, then expression decreased and was close to the minimum values characteristic of control tFIBb samples.

The expression of some genes did not change significantly in all samples (*LusPILS2-C*, and *D*) or was increased in fibers at an early stage of TCW formation (tFIBa) (*LusPIN5-B*, *LusPILS1-C*, *D*, *LusPILS6-A*, *B*, and *LusPILS7-B*), but did not differ significantly in fibers during gravistimulation (data not shown). Figure 4 shows the gene expression, the maxi-

imum value of which was observed during gravistimulation. *LusPIN5-A* had a maximum expression level in PUL fibers at 96 hours after gravistimulation. *LusPILS3* had an increased expression level in OPP samples after 8 hours, while the peak of expression was also observed in PUL samples, but only 24 hours after the beginning of gravistimulation (see Fig. 4).

Three of the 22 genes showed an increased level of expression only during gravistimulation, and especially in PUL samples after 24 hours (*LusPIN8-B*), or 8 hours from the beginning of gravistimulation (*LusPILS1-B*, *LusPILS7-A*). The expression of *LusPIN8-B* decreased in PUL samples

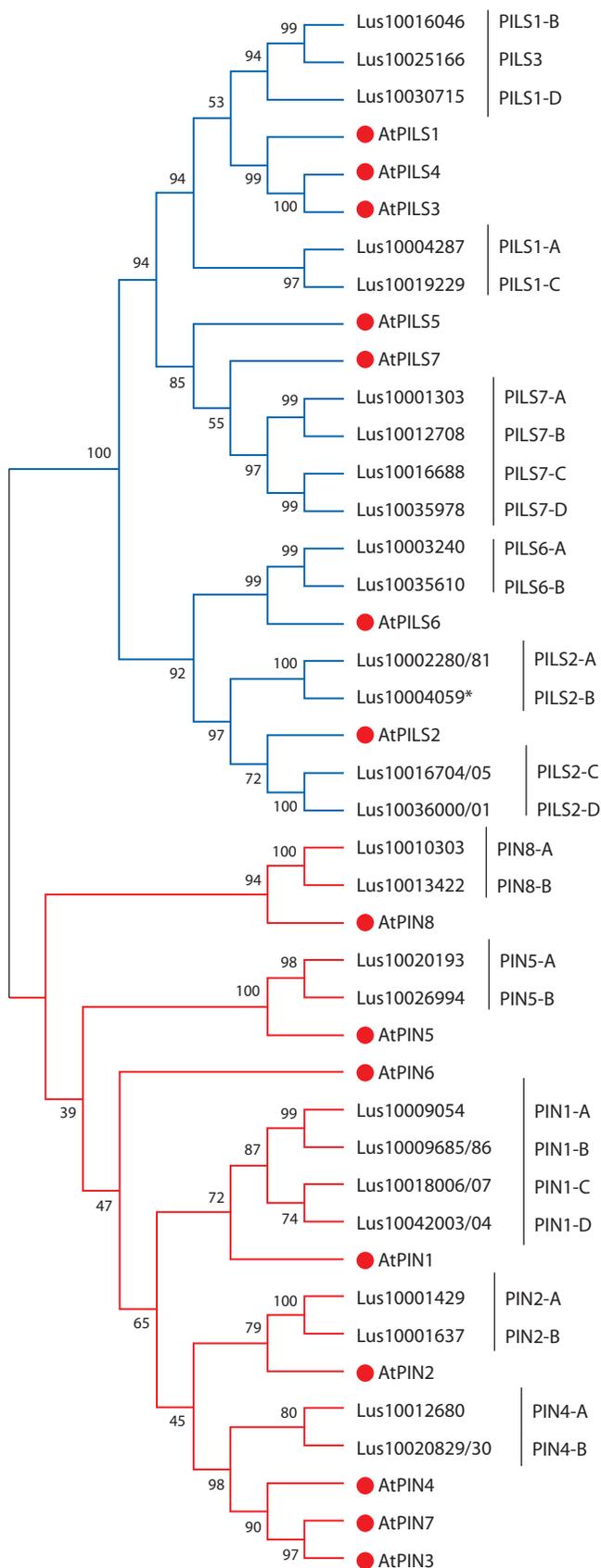


Fig. 1. Phylogenetic tree for amino acid sequences PF03547 in *A. thaliana* and *L. usitatissimum*.

Maximum Likelihood method, model Le_Gascuel_2008 (LG+G). Bootstrap support 1000. Performed in the MEGA 7 program. The red marker indicates the sequences of *A. thaliana*.

after 96 hours, but at the same time increased in OPP (Fig. 5). The most contrasting expression between PUL and OPP was demonstrated for *LusPILS7-A*, which was almost leveled after 96 hours.

Identification and expression of *LusAUX1/LAX*

Auxin influx carriers *LusAUX1/LAX* have a conservative PF01490 domain (Transmembrane amino acid transporter protein). In *A. thaliana*, the *AUX1/LAX* family is represented by four highly conserved genes called *AUX1*, *LAX1*, *LAX2*, and *LAX3*, which encode proteins similar to amino acid carriers (Young et al., 1999). In total, 82 genes of flax with PF01490 are represented in the Phytozome database. Of these, eight encode auxin transporters and correspond to three *A. thaliana* orthologous genes (*AUX1*, *AUX2*, and *LAX3*). All genes in this group encoded proteins close in molecular weight and isoelectric point; 10 transmembrane domains were predicted for all proteins (Table 2).

The genes encoding *LusLAX2* (A and B) were highly expressed in intrusively growing fibers (iFIBa), while their expression dropped sharply in the fibers forming TCW and remained low during gravistimulation (Fig. 6).

The expression dynamics of *LusLAX3* (A and B) were absolutely different from *LusLAX2*, while the expression dynamics between the two paralogs were identical, as in the case of *LusLAX2*. *LusLAX3* had the maximum expression level in the fibers forming TCW at a late stage (tFIBb). During the gravity response, the expression level of these genes dropped sharply (Fig. 7).

LusAUX1 genes had a relatively high level of expression in all samples, while four paralogs showed a clear division into two groups according to the expression patterns: with the maximum level in the fibers forming TCW (tFIBa, *LusAUX1-A*, and *D*), and with the maximum level of expression in the fibers of gravistimulated plants (OPP, 24 hours) (*LusAUX1-B*, and *C*) (Fig. 8).

Identification and expression of *LusABCB*

According to the Phytozome database, 206 ABC transporter genes are present in the flax genome, of which 32 genes belong to group B. To analyze the expression, we selected orthologs of the listed *A. thaliana* genes in flax. 25 genes corresponding to 5 Arabidopsis orthologs were identified: *ABCB1* (2 flax genes), *ABCB4* (4 flax genes), *ABCB15* (9 flax genes), *ABCB19* (8 flax genes), and *ABCB20* (2 flax genes). The *Lus10011977* sequence was partially corrected, and *Lus10036616* and *Lus10036617* were combined into one sequence (Table 3). Of the 24 genes, 4 (*ABCB4* and 3 isoforms of *ABCB15*) had a low level of expression or were not expressed.

The remaining 21 genes had different levels and patterns of expression. Thus, *LusABCB20* (A and B) had a high level of expression in growing fibers, and at an early stage of TCW formation, in mature fibers, their expression decreased and almost did not change during gravistimulation. Two of the four isoforms of *LusABCB4* had a maximum expression level in growing fibers, while the expression level itself was low, and the third isoform had a peak expression in fibers at an early stage of the formation of TCW. The most diverse expression patterns were characteristic of *LusABCB19*, which also had

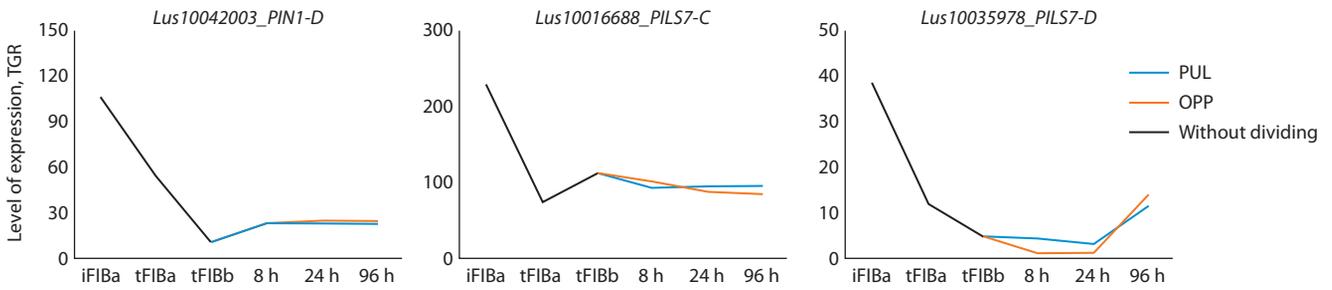


Fig. 2. Expression of *LusPIN/PILS* with an increased level of expression only in intrusively growing fibers.

Here and in Fig. 3–9: iFIBa – intrusively growing fibers; tFIBa – the early stage of tertiary cell wall (TCW) formation; tFIBb – the late stage of TCW formation. 8, 24, 96 hours – the time of fixation of the samples after the stem inclination; TGR – total gene reads; PUL – pulling side; OPP – opposite to the PUL-side. Without dividing – the plants were not subjected to gravistimulation, and the stem was not divided into PUL and OPP.

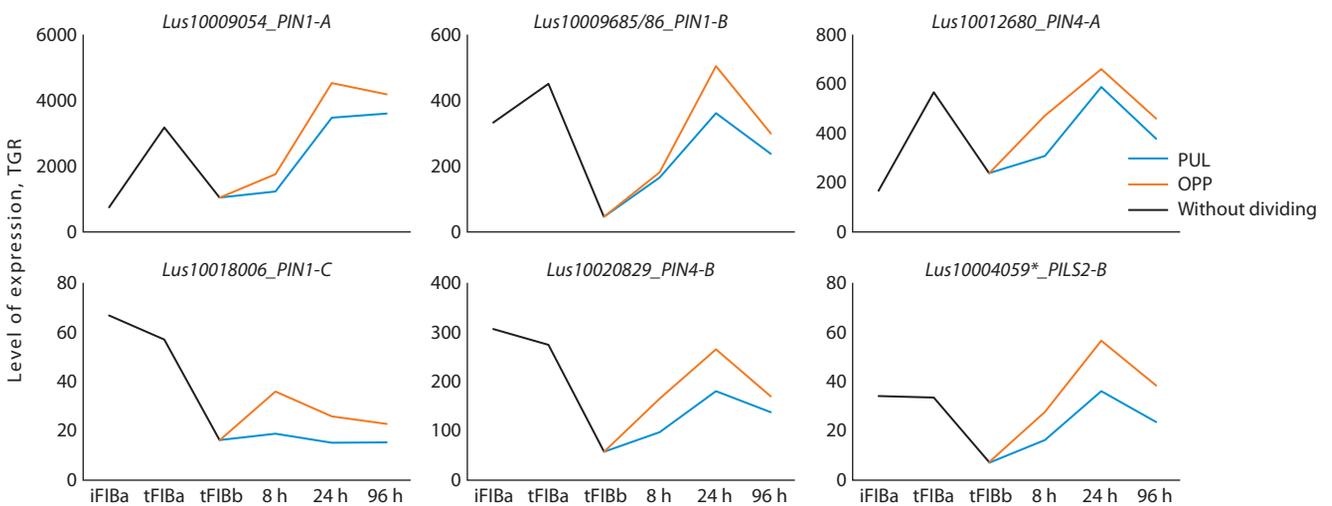


Fig. 3. Expression of *LusPIN1-A, B, C*, *LusPIN4-A, B*, *LusPILS2-B* in flax fibers under normal conditions and in gravity response.

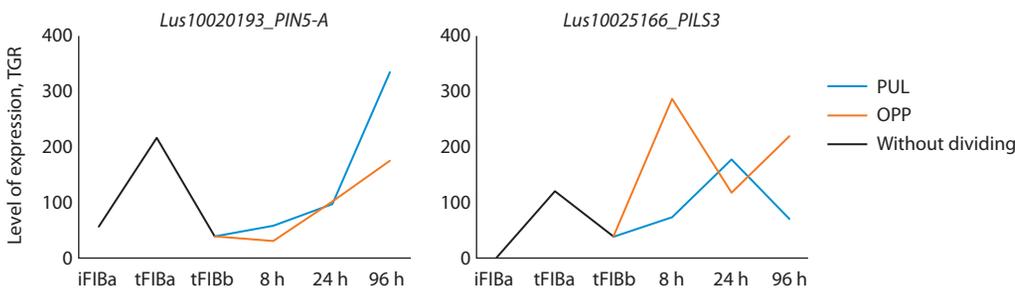


Fig. 4. Expression of *LusPIN5-A* and *LusPILS3* in flax fibers under normal conditions and in gravity response.

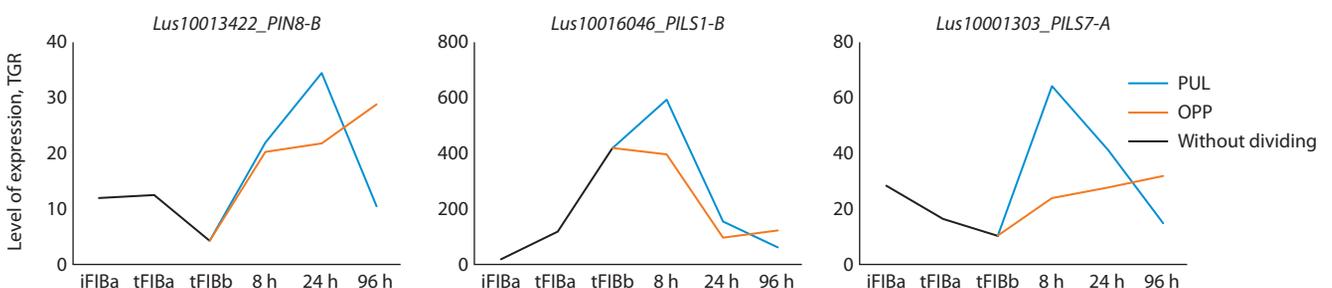


Fig. 5. Expression of *LusPIN8-B*, *LusPILS1-B*, *LusPILS7-A* in flax fibers under normal conditions and in gravity response.

Table 2. List and some characteristics of *LusAUX1/LAX* sequences

<i>LUS</i>	<i>AT</i>	<i>LUS</i> Name	AA	Mw, kDa	pI	TMH_ <i>LUS</i>
Lus10028278	AT2G38120	AUX1-A	628	70.0	7.2	10
Lus10002498	AT2G38120	AUX1-B	497	55.7	8.7	10
Lus10004831	AT2G38120	AUX1-C	487	54.7	8.7	10
Lus10040212	AT2G38120	AUX1-D	486	54.7	8.9	10
Lus10025057	AT2G21050	LAX2-A	497	55.6	8.5	10
Lus10034488	AT2G21050	LAX2-B	497	55.5	8.6	10
Lus10028078	AT1G77690	LAX3-A	477	53.7	8.7	10
Lus10025628	AT1G77690	LAX3-B	478	53.8	8.6	10

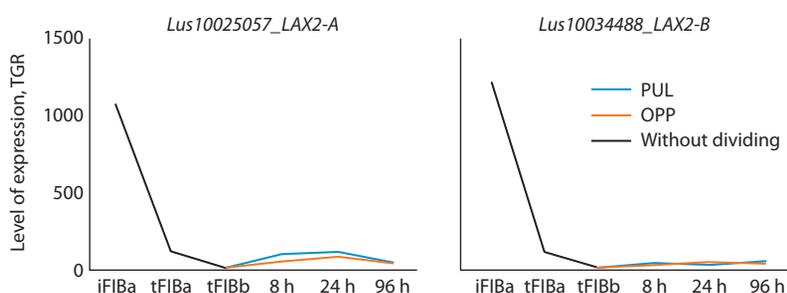


Fig. 6. Expression of *LusLAX2-A, B* in fibers under normal conditions and in gravity response.

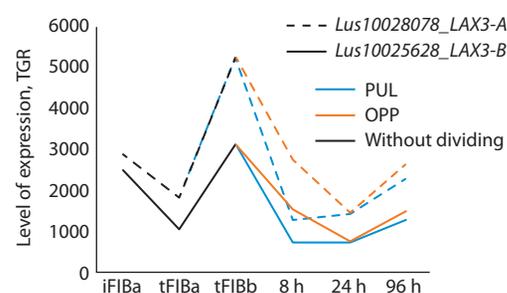


Fig. 7. Expression of *LusLAX3-A, B* in flax fibers under normal conditions and in gravity response.

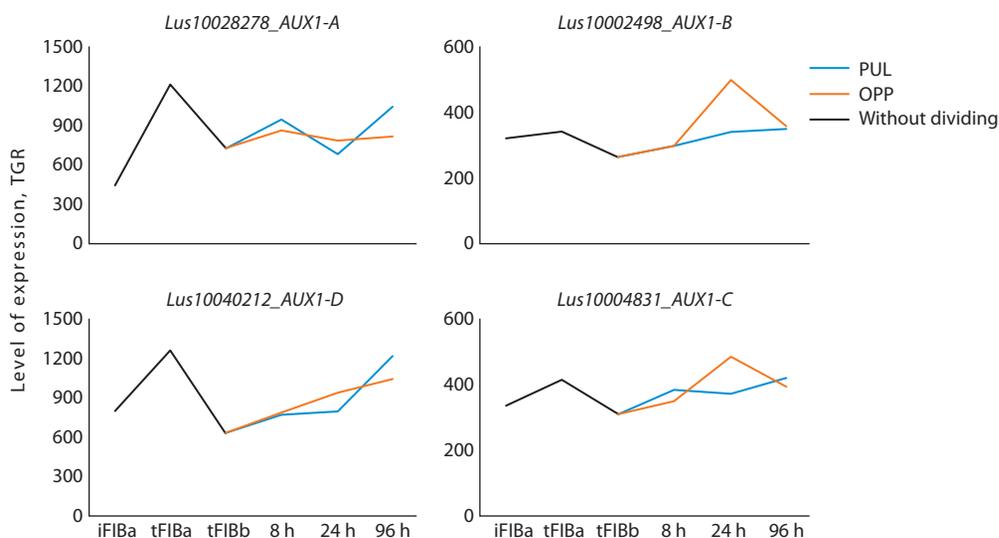


Fig. 8. Expression of *LusAUX1-A, B, C, D* in flax fibers under normal conditions and in gravity response.

the largest number of expressed isoforms (8 genes) (data are not provided).

We selected *LusABCB* as having the maximum difference in expression between PUL and OPP samples. Among the 6 genes, 4 genes had increased expression in the fibers forming TCW; the expression of these genes decreased in more mature fibers but increased in the fibers of gravistimulated plants,

especially in OPP samples (significantly for *ABCB15-B* and *ABCB19-B*) (Fig. 9). A low level of expression was observed for *LusABCB15-B*, but the gene was specifically activated during gravistimulation and was practically not expressed in other samples. The expression of this gene is almost five times higher in OPP-side fibers compared to PUL (8, 24 hours) (see Fig. 9).

Table 3. List and some characteristics of LusABCB sequences

LUS	AT	LUS Name	AA	Mw, kDa	pI	TMH_LUS
Lus10024162	AT3G28345	ABCB15-E	1246	135.7	8.5	9
Lus10033470	AT3G28860	ABCB19-F	1511	164.8	8.8	12
Lus10014427	AT2G36910	ABCB1-A	1210	132.1	8.8	9
Lus10015595	AT3G28860	ABCB19-G	1268	139.2	7.0	11
Lus10041565	AT3G28345	ABCB15-B	1244	135.2	7.8	10
Lus10035834	AT3G28345	ABCB15-H	1287	139.8	8.9	7
Lus10030674	AT3G28860	ABCB19-B	1254	136.6	8.0	8
Lus10005249	AT3G28860	ABCB19-A	1254	136.5	8.0	8
Lus10012959	AT3G28345	ABCB15-D	1237	135.4	8.6	9
Lus10032911	AT3G28860	ABCB19-H	1269	138.9	7.2	9
Lus10023437	AT3G55320	ABCB20-A	1406	155.4	6.3	12
Lus10023929	AT2G36910	ABCB1-B	1338	145.9	7.3	12
Lus10011977*	AT2G47000	ABCB4-D	1072	117.4	8.3	2
Lus10038050	AT2G47000	ABCB4-B	1216	131.1	8.4	9
Lus10013178*	AT3G28860	ABCB19-D	2432	266.4	6.7	10
Lus10039533	AT3G28345	ABCB15-F	1250	136.4	8.2	9
Lus10008139	AT3G28860	ABCB19-C	1249	136.1	7.4	11
Lus10040315	AT3G55320	ABCB20-B	1395	154.1	6.4	12
Lus10036616/17	AT3G28345	ABCB15-G	1267	137.8	8.9	9
Lus10039458	AT3G28345	ABCB15-A	1270	137.7	8.0	9
Lus10009989	AT2G47000	ABCB4-C	1272	136.9	7.7	9
Lus10020905	AT3G28860	ABCB19-E	1504	163.6	8.8	12
Lus10005839	AT3G28345	ABCB15-C	1261	136.8	8.3	10
Lus10004520*	AT2G47000	ABCB4-A	826	89.2	8.2	7

* The sequence may be incorrect.

Discussion

The action of auxin as a switch is closely related to the presence of local maximums and minimums formed in tissues (Adamowski, Friml, 2015). They are created, maintained, and modulated by intercellular auxin transfer, a plant-specific process. This process, called polar auxin transport, depends on the action of representatives of at least three auxin carrier families: PIN-FORMED, AUX1/LAX, and ABCB (Geisler et al., 2017).

In this study, the main auxin carrier genes were identified in flax plants: 12 *LusPINs*, 15 *LusPILS*, 8 *LusAUX1/LAX*, and 24 *LusABCB*. A comparative analysis of the expression of these genes in flax phloem fibers at different stages of development revealed increased expression of some of them at the stage of intrusive growth (*LusLAX2* (A, B), *LusPIN1-D*, *LusPILS7* (C, D)), at the early stage of TCW formation (*LusAUX1* (A, D), *LusABCB1-A, B*, *LusABCB15-A*, *LusPIN1-*

A, B, *LusPIN4-A*, *LusPIN5-A*), and at the late stage of TCW formation (*LusLAX3* (A, B)).

As known, all auxin transporters can be simplistically divided into three groups: responsible for the influx of auxin into the cell, outflow from the cell and intracellular transport. Two types of carriers participate in the outflow of auxin from the cell: PIN proteins and ABCB carriers (Zažímalová et al., 2010). PIN protein classification is usually based on phylogenetic relationships, subcellular localization, and the length of hydrophilic loop domains. From this point of view, members of the PIN protein family are usually grouped into three types: (1) canonical (PIN1, 2, 3, 4 and 7 – for *A. thaliana*), localized on the plasmalemma (PM), which mediate the intercellular flow of auxin; (2) non-canonical (PIN5 and 8 – for *A. thaliana*), which are localized on the EPR membrane and mediate auxin exchange between the cytosol and the EPR, contributing to intracellular auxin homeostasis; and (3) double, PM- and

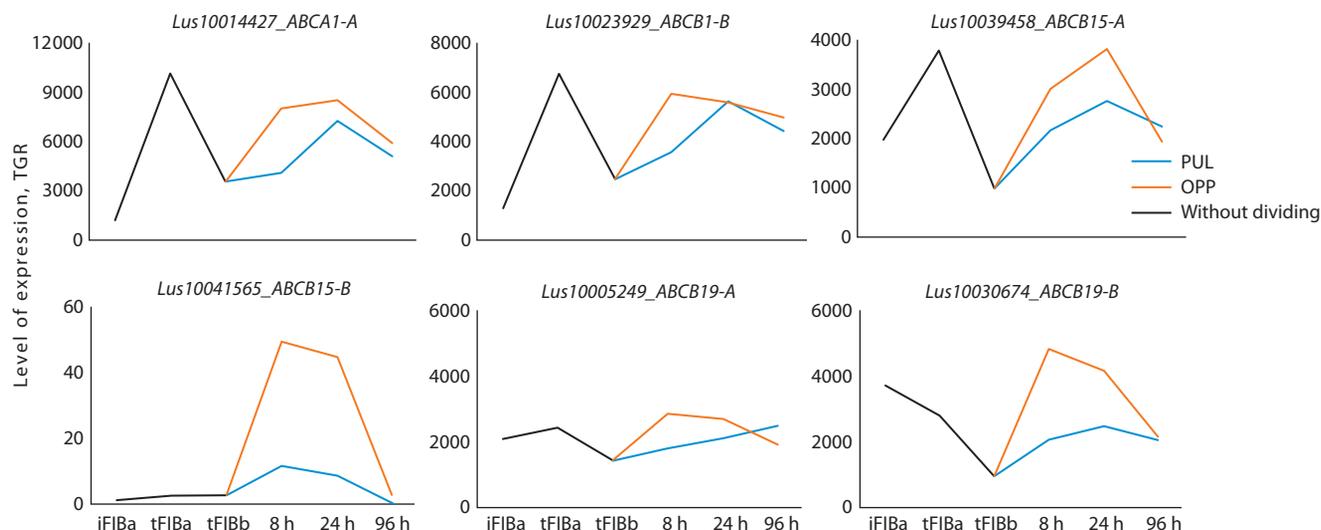


Fig. 9. Expression of some *LusABC1*, *15*, and *19* gene isoforms in fibers under normal conditions and in gravity response.

EPR-localized, PINs, such as PIN6 in *A. thaliana*, with unclear function (Zhang et al., 2020). According to the results obtained in this study, the following trend was observed for PIN carriers: the genes of canonical *LusPINs* were highly expressed in fibers at the early stage of TCW formation (see Fig. 3) and in fibers of gravistimulated plants, and their increased expression was observed in OPP samples (see Fig. 3), while increased expression of non-canonical PIN genes was observed in PUL samples (see Fig. 4, 5), which gives us the opportunity to assume a redistribution of auxin content in fibers on different stem sides during graviresponse. In confirmation of this assumption, a similar trend was also revealed in relation to ABCB carriers (see Fig. 9): a higher expression of *LusABC1*, *15*, and *19* in OPP samples was shown. ABCB transporters carry out transport due to the direct binding of ATP and the energy that is released during ATP hydrolysis and can function when chemiosmotic gradients decrease or when auxin must move against the gradient (Zažimalová et al., 2010). Arabidopsis contains 21 full-sized *ABCB* genes (Kang et al., 2011), but only for four isoforms (*ABCB1*, *ABCB4*, *ABCB19*, and *ABCB21*) reliable data concerning auxin transport were obtained; for *ABCB1* and *ABCB19*, data on coordinated action under gravitropism have been demonstrated (Geisler et al., 2017). It has also recently been shown that the pairs *ABCB1/19* and *ABCB6/20* represent the main ABCB auxin carriers over long distances through the vascular system in Arabidopsis (Jeness et al., 2022). There is an assumption that *ABCB14* and *ABCB15* are involved in auxin transport during stem lignification (Kaneda et al., 2011). It should be noted that the expression of genes for the ABCB transporter in Arabidopsis seedlings was studied in the roots, hypocotyl, and apex of the shoot (Geisler et al., 2017). It was assumed that ABCBs can play the role of the main auxin carriers: they are uniformly localized on PM, are usually found in various plant species, and persist stably regardless of internal and external signals. On the contrary, PINs are asymmetrically localized and dynamically distributed in response to endogenous and exogenous signals (Cho M., Cho H.T., 2013).

It should be noted that in this study, two paralogs homologous to the *AtPIN3/4/7* clade were identified (see Fig. 1), which we annotated by the closest homologue as *LusPIN4* (A and B). *PIN3* is known to provide lateral auxin transport (Friml et al., 2002; Rakusová et al., 2019); high expression of *PIN3* is shown in mature rami fibers with a thickened TCW (Bao et al., 2019), and during tension wood formation (Gerttula et al., 2015). In this study, the *LusPIN4* genes (A and B) significantly increased expression during gravistimulation (24 hours) (see Fig. 3), but expression in the fibers of OPP samples was slightly higher compared to PUL samples. A similar trend was shown for gene expression and membrane localization of *AtPIN3* and *AtPIN4* during the hypocotyl apical hook formation, where pulling and opposite sides are also observed. The authors suggested that an increase in the content of *PIN3* and *PIN4* in the cell membrane on the opposite side is a decisive factor for the local auxin maximum formation (Zhu et al., 2019). It should be noted that we investigated the part of the stem of mature plants where the fibers do not grow by elongation, where the second (located below) curvature is formed. We have previously shown that when the upper part of the stem is removed (where the upper curvature is formed), plants implement the gravitropic reaction no less successfully (Ibragimova et al., 2017).

In the current work, it was shown that *AUX1/LAX* genes responsible for auxin influx into the cell increased expression during gravireaction to the level of expression in fibers at an early stage of TCW formation (*LusAUX1-D* for PUL), but there was no significant difference between OPP and PUL (see Fig. 8). However, a very interesting fact is that the highly expressed genes *Lus10028078* and *Lus10025628* (*ATIG77690 – LAX3*) had a single maximum in the control samples at the late stage of TCW formation (see Fig. 7). These data are consistent with the high expression of similar genes observed in mature rami fibers (Bao et al., 2019). With gravity response, the expression of these genes decreased and increased slightly towards the end of the reaction, approaching the values in the phase of the beginning of the formation

of TCW (see Fig. 7). For *LAX2*, another effect was observed: the maximum expression in the fibers occurred in the phase of intrusive growth, but, as for *LAX3* and *AUX1*, with gravity response, the expression values were again close to values at the early stage of TCW formation (significantly for the *LusLAX2-A* gene on the PUL side) (see Fig. 6).

The results of our study show that the increased expression of *LusPILS* in fibers quite often occurred during gravity response, both in the fibers of the PUL and OPP sides (see Fig. 4, 5), which suggests the presence of a relationship between gravity reaction and intracellular redistribution of auxin in general. At the same time, in some cases, a significant difference was observed in the fibers on different sides of the stem (*LusPILS3* and *LusPILS7-A*) (see Fig. 4, 5). It should be noted that in *LusPILS7-C* and *D*, high expression was observed only at the stage of intrusive fiber growth (see Fig. 2). *PILS* transporters have been identified *in silico* as a putative family of auxin transport mediators; it has been shown that *PILS*, including *PILS3* and *PILS7*, regulate the accumulation of auxin in the cell by retaining the added auxin (Barbez et al., 2012), which may also occur in the case of gravistimulation.

Thus, it was shown that during the gravitropic reaction, the genes encoding transporters responsible for auxin outflow from the cell (*PIN* and *ABC*) and intracellular transport responsible for auxin homeostasis (non-canonical *PIN5*, *PIN8* and *PILS*) were most significantly expressed. The expression level of these genes often approached the values that were characteristic of the stage at the beginning of TCW formation, when biosynthetic processes proceeded more intensively compared to those in mature fibers. It was shown that the expression of transporter genes responsible for the influx of auxin into cells (*LusAUX1-D*) also increased during gravitropism. The differential expression of the genes of the IAA carriers in the fibers located on different sides of the stem was revealed: the difference was observed due to the expression of genes, the products of which are responsible for intracellular transport (*LusPILS3*, *LusPILS7-A*) and auxin outflow (*LusABC15-B*, *LusABC19-B*). Increased expression of *PIN* genes and *ABC* genes was more typical for the fibers of the OPP side of the stem.

Conclusion

In this study, the main auxin carrier genes were identified in flax plants: 8 *LusAUX1/LAX*, 12 *LusPIN-FORMED*, 15 *LusPIN-LIKES*, and 24 *LusABC*. The differential expression of the genes of the IAA transporters at different stages of development and in the fibers located on different stem sides during gravity response was revealed. We assume that the realization of this reaction may be associated with an asymmetric redistribution of auxin, mainly due to auxin intracellular transporters and transporters responsible for its outflow from the cell, an increase in the gene expression of which we observed during gravireaction. Further studies are required to clarify the mechanisms of auxin's participation in the implementation of a gravity response not associated with growth by elongation, along with the participation of other hormones in it.

References

- Adamowski M., Friml J. PIN-dependent auxin transport: action, regulation, and evolution. *Plant Cell*. 2015;27(1):20-32. DOI 10.1105/tpc.114.134874
- Bao Y., Huang X., Rehman M., Wang Y., Wang B., Peng D. Identification and expression analysis of the PIN and AUX/LAX gene families in ramie (*Boehmeria nivea* L. Gaud). *Agronomy*. 2019;9:435. DOI 10.3390/agronomy9080435
- Barbez E., Kubeš M., Rolčík J., Béziat C., Pencík A., Wang B., Rosquete M.R., Zhu J., Dobrev P.I., Lee Y., Zažimalová E., Petrášek J., Geisler M., Friml J., Kleine-Vehn J. A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. *Nature*. 2012;485(7396):119-122. DOI 10.1038/nature11001
- Cho M., Cho H.T. The function of ABCB transporters in auxin transport. *Plant Signal. Behav.* 2013;8(2):e22990. DOI 10.4161/psb.22990
- Evans M.L. Gravitropism: interaction of sensitivity modulation and effector redistribution. *Plant Physiol.* 1991;95(1):1-5. DOI 10.1104/pp.95.1.1
- Friml J., Wiśniewska J., Benkova E., Mendgen K., Palme K. Lateral relocation of auxin efflux regulator PIN3 mediates tropism in *Arabidopsis*. *Nature*. 2002;415(6873):806-809. DOI 10.1038/415806a
- Geisler M., Aryal B., Donato M., Hao P.A. Critical view on ABC transporters and their interacting partners in auxin transport. *Plant Cell Physiol.* 2017;58(10):1601-1614. DOI 10.1093/pcp/pcx104
- Gerttula S., Zinkgraf M., Muday G., Lewis D., Ibatullin F., Brumer H., Hart F., Mansfield S., Filkov V., Groover A. Transcriptional and hormonal regulation of gravitropism of woody stems in *Populus*. *Plant Cell*. 2015;27(10):2800-2813. DOI 10.1105/tpc.15.00531
- Gorshkova T.A., Sal'nikov V.V., Chemikosova S.B., Ageeva M.V., Pavlencheva N.V. The snap point: a transition point in *Linum usitatissimum* bast fiber development. *Ind. Crops Prod.* 2003;18(3):213-221. DOI 10.1016/S0926-6690(03)00043-8
- Harrison M.A., Pickard B.G. Auxin asymmetry during gravitropism by tomato hypocotyls. *Plant Physiol.* 1989;89(2):652-657. DOI 10.1104/pp.89.2.652
- Haygreen J.G., Bowyer J.L. Forest Products and Wood Science. Wiley-Blackwell, 1996
- Hellgren J.M., Olofsson K., Sundberg B. Patterns of auxin distribution during gravitational induction of reaction wood in poplar and pine. *Plant Physiol.* 2004;135(1):212-220. DOI 10.1104/pp.104.038927
- Ibragimova N.N., Ageeva M.V., Gorshkova T.A. Development of gravitropic response: unusual behavior of flax phloem G-fibers. *Protoplasma*. 2017;254(2):749-762. DOI 10.1007/s00709-016-0985-8
- Ibragimova N., Mokshina N., Ageeva M., Gurjanov O., Mikshina P. Rearrangement of the cellulose-enriched cell wall in flax phloem fibers over the course of the gravitropic reaction. *Int. J. Mol. Sci.* 2020;21(15):5322. DOI 10.3390/ijms21155322
- Jenness M.K., Tayengwa R., Bate G.A., Tapken W., Zhang Y., Pang C., Murphy A.S. Loss of multiple ABCB auxin transporters recapitulates the major twisted dwarf 1 phenotypes in *Arabidopsis thaliana*. *Front. Plant Sci.* 2022;13:840260. DOI 10.3389/fpls.2022.840260
- Jourez B., Riboux A., Leclercq A. Anatomical characteristics of tension wood and opposite wood in young inclined stems of poplar (*Populus euramericana* cv 'Ghoy'). *LAWA J.* 2001;22(2):133-157. DOI 10.1163/22941932-90000274
- Kaneda M., Schuetz M., Lin B.S., Chanis C., Hamberger B., Western T.L., Ehltling J., Samuelset A.L. ABC transporters coordinately expressed during lignifications of *Arabidopsis* stems include a set of ABCBs associated with auxin transport. *J. Exp. Bot.* 2011;62(6):2063-2077. DOI 10.1093/jxb/erq416
- Kang J., Park J., Choi H., Burla B., Kretschmar T., Lee Y., Martinioia E. Plant ABC transporters. *Arabidopsis Book*. 2011;9:e0153. DOI 10.1199/tab.0153
- Li Y., Hagen G., Guilfoyle T.J. An auxin-responsive promoter is differentially induced by auxin gradients during tropisms. *Plant Cell*. 1991;3(11):1167-1175. DOI 10.1105/tpc.3.11.1167

- Manna M., Rengasamy B., Ambasht N.K., Sinha A.K. Characterization and expression profiling of PIN auxin efflux transporters reveal their role in developmental and abiotic stress conditions in rice. *Front. Plant Sci.* 2022;13:1059559. DOI 10.3389/fpls.2022.1059559
- Mokshina N., Gorshkov O., Takasaki H., Onodera H., Sakamoto S., Gorshkova T., Mitsuda N. FIBexDB: a new online transcriptome platform to analyze development of plant cellulosic fibers. *New Phytol.* 2021;231(2):512-515. DOI 10.1111/nph.17405
- Rakusová H., Han H., Valošek P., Friml J. Genetic screen for factors mediating PIN polarization in gravistimulated *Arabidopsis thaliana* hypocotyls. *Plant J.* 2019;98(6):1048-1059. DOI 10.1111/tpj.14301
- Swarup R., Péret B. AUX/LAX family of auxin influx carriers – an overview. *Front. Plant Sci.* 2012;3:225. DOI 10.3389/fpls.2012.00225
- Timell T.E. The chemical composition of tension wood. *Svensk Papperstidning.* 1969;72:173-181
- Young G.B., Jack D.L., Smith D.W., Saier M.H., Jr. The amino acid/auxin:proton symport permease family. *Biochim. Biophys. Acta.* 1999;1415(2):306-322. DOI 10.1016/s0005-2736(98)00196-5
- Zažimalová E., Murphy A.S., Yang H., Hoyerová K., Hošek P. Auxin transporters – why so many? *Cold Spring Harb. Perspect. Biol.* 2010;2(3):a001552. DOI 10.1101/cshperspect.a001552
- Zhang Y., Hartinger C., Wang X., Friml J. Directional auxin fluxes in plants by intramolecular domain-domain coevolution of PIN auxin transporters. *New Phytol.* 2020;227(5):1406-1416. DOI 10.1111/nph.16629
- Zhu Q., Gallemí M., Pospíšil J., Žádníková P., Strnad M., Benková E. Root gravity response module guides differential growth determining both root bending and apical hook formation in *Arabidopsis*. *Development.* 2019;146(17):dev175919. DOI 10.1242/dev.175919.

ORCID

N.E. Mokshina orcid.org/0000-0002-6434-7404

Acknowledgements. The work was carried out with the financial support of the RSF grant No. 23-24-00612.

Conflict of interest. The authors declare no conflict of interest.

Received July 26, 2023. Revised November 15, 2023. Accepted November 16, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Ascorbate-glutathione cycle in wheat and rice seedlings under anoxia and subsequent reaeration

V.V. Yemelyanov^{1, 2} , E.G. Prikaziuk^{2, 3}, V.V. Lastochkin², O.M. Aresheva², T.V. Chirkova²

¹ Department of Genetics and Biotechnology, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia

² Department of Plant Physiology and Biochemistry, Faculty of Biology, Saint Petersburg State University, St. Petersburg, Russia

³ Department of Water Resources, ITC Faculty of Geo-Information Science and Earth Observation, University of Twente, Enschede, the Netherlands

 bootika@mail.ru

Abstract. The most important part of the plant antioxidant system is the ascorbate-glutathione cycle (AGC), the activity of which is observed upon exposure to a range of stressors, including lack of O₂, and oxidative stress occurring immediately after the restoration of oxygen access, hereafter termed reaeration or post-anoxia. The operation of the AGC (enzymes and low-molecular components) in wheat (*Triticum aestivum*, cv. Leningradka, non-resistant to hypoxia) and rice (*Oryza sativa*, cv. Liman, resistant) seedlings after 24 h anoxia and 1 h or 24 h reaeration was studied. Significant accumulation of oxidized forms of ascorbate and glutathione was revealed in the non-resistant plant (wheat) after 24 h of anoxia and reaeration, indicating the development of oxidative stress. In the resistant plant (rice), reduced forms of these antioxidants prevailed both in normoxia and under stress, which may indicate their intensive reduction. In wheat, the activities of ascorbate peroxidase and dehydroascorbate reductase in shoots, and monodehydroascorbate reductase and glutathione reductase in roots decreased under anoxia and reaeration. The activity of antioxidant enzymes was maintained in rice under lack of oxygen (ascorbate peroxidase, glutathione reductase) and increased during post-anoxia (AGC reductases). Anoxia stimulated accumulation of mRNA of the organellar ascorbate peroxidase genes *OsAPX3*, *OsAPX5* in shoots, and *OsAPX3-5* and *OsAPX7* in roots. At post-anoxia, the contribution of the *OsAPX1* and *OsAPX2* genes encoding the cytosolic forms of the enzyme increased in the whole plant, and so did that of the *OsAPX8* gene for the plastid form of the enzyme. The accumulation of mRNA of the genes *OsMDAR2* and *OsMDAR4* encoding peroxisomal and cytosolic monodehydroascorbate reductase as well as the *OsGR2* and *OsGR3* for cytosolic and organellar glutathione reductase was activated during reaeration in shoots and roots. In most cases, O₂ deficiency activated the genes encoding the peroxisomal, plastid, and mitochondrial forms of the enzymes, and upon reaeration, an enhanced activity of the genes encoding the cytoplasmic forms was observed. Taken together, the inactivation of AGC enzymes was revealed in wheat seedlings during anoxia and subsequent reaeration, which disrupted the effective operation of the cycle and triggered the accumulation of oxidized forms of ascorbate and glutathione. In rice, anoxia led to the maintenance of the activity of AGC enzymes, and reaeration stimulated it, including at the level of gene expression, which ensured the effective operation of AGC.

Key words: anoxia; reaeration; oxidative stress; ascorbate; glutathione; ascorbate-glutathione cycle; wheat; rice.

For citation: Yemelyanov V.V., Prikaziuk E.G., Lastochkin V.V., Aresheva O.M., Chirkova T.V. Ascorbate-glutathione cycle in wheat and rice seedlings under anoxia and subsequent reaeration. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(1):44-54. DOI 10.18699/vjgb-24-06

Аскорбат-глутатионовый цикл в проростках пшеницы и риса при аноксии и последующей реэрации

В.В. Емельянов^{1, 2} , Е.Г. Приказюк^{2, 3}, В.В. Ласточкин², О.М. Арешева², Т.В. Чиркова²

¹ Кафедра генетики и биотехнологии, биологический факультет, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

² Кафедра физиологии и биохимии растений, биологический факультет, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

³ Кафедра водных ресурсов, факультет геоинформатики и наблюдения за Землей, Университет Твенте, Энсхеде, Нидерланды

 bootika@mail.ru

Аннотация. Важной частью антиоксидантной системы растений является аскорбат-глутатионовый цикл (АГЦ), функционирование которого регистрируется при действии разнообразных стрессоров, в том числе дефицита и/или отсутствия O₂, а также окислительного стресса, возникающего сразу после восстановления доступа кислорода (реэрация, или постаноксия). Показана значительная аккумуляция окисленных форм аскорбата и глутатиона в проростках пшеницы (*Triticum aestivum*, сорт Лeningradka, неустойчивое растение) при действии на него 24-часовой аноксии и реэрации, что свидетельствует о развитии окислительного стресса. У риса (*Oryza sativa*, сорт Лиман, устойчивое растение) преобладали восстановленные формы данных антиоксидантов как в контроле, так и при стрессе, что может указывать на их интенсивное восстановление. У пшеницы активности аскорбат-

пероксидазы и дегидроаскорбатредуктазы в побегах, а также монодегидроаскорбатредуктазы и глутатионредуктазы в корнях снижались под действием аноксии и реаэрации. У риса активность антиоксидантных ферментов сохранялась в отсутствие кислорода (аскорбатпероксидаза, глутатионредуктаза) и возрастала при постаноксии (редуктазы АГЦ). Аноксия стимулировала накопление мРНК генов органелльных форм аскорбатпероксидазы *OsAPX3*, *OsAPX5* в побегах и *OsAPX3-5*, *OsAPX7* в корнях проростков риса. При постаноксии во всем растении возрастал вклад генов цитоплазматических форм фермента – *OsAPX1*, *OsAPX2*, а также пластидной *OsAPX8*. При реаэрации накапливались транскрипты пероксисомной и цитоплазматической монодегидроаскорбатредуктазы *OsMDAR2* и *OsMDAR4*, цитозольной и органелльной глутатионредуктазы *OsGR2* и *OsGR3*. В большинстве случаев недостаток O_2 индуцировал активность генов, кодирующих пероксисомные, пластидные и митохондриальные формы ферментов, а реаэрация усиливала работу генов, кодирующих цитоплазматические формы. При действии аноксии и последующей реаэрации выявлена инактивация ферментов АГЦ в проростках пшеницы, что нарушало эффективную работу цикла и запускало аккумуляцию окисленных форм аскорбата и глутатиона. У риса аноксия приводила к сохранению активности ферментов АГЦ, а реаэрация ее стимулировала, в том числе на уровне экспрессии генов, что обеспечивало эффективное функционирование аскорбат-глутатионового цикла. Ключевые слова: аноксия; реаэрация; окислительный стресс; аскорбиновая кислота; глутатион; аскорбат-глутатионовый цикл; пшеница; рис.

Introduction

Plants have a multi-level system of protection against the damaging effects of reactive oxygen species (ROS), which accumulated in response to changes in environmental conditions, developmental stages, the action of hormones, etc. (Halliwell, 2006; Foyer, Noctor, 2009; Shikov et al., 2021). An important part of the antioxidant system is the ascorbate-glutathione cycle (AGC, Foyer–Halliwell–Asada pathway), which ensures the effective reduction of low-molecular-weight antioxidants – ascorbate and glutathione. Ascorbic acid (AsA, vitamin C) is a key water-soluble antioxidant present in all cell compartments, including the apoplast (Noctor, Foyer, 1998). AsA neutralizes most ROS and reduces tocopherols and epoxy-carotenoids of the violaxanthin cycle. When AsA is oxidized, either a short-lived radical, monodehydroascorbic acid (MDA), or a non-radical oxidized form, dehydroascorbic acid (DHA), is formed. Two molecules of MDA can undergo a disproportionation reaction to form one molecule of reduced AsA and one molecule of DHA.

Glutathione, another low-molecular-weight AGC antioxidant, is a γ -Glu-Cys-Gly tripeptide. It is present throughout the cell, but in the apoplast, only in trace amounts (Noctor, Foyer, 1998; Gill, Tuteja, 2010). Glutathione neutralizes reactive oxygen and nitrogen species, free radicals and fatty acid peroxides, participates in the neutralization of methylglyoxal, xenobiotics and heavy metals, and reduces DHA and sulfhydryl groups (Hasanuzzaman et al., 2017). It is involved in various physiological processes, including redox regulation, signal transduction, conjugation and transport of metabolites, regulation of plant growth and development (Gill, Tuteja, 2010).

The key enzyme of the AGC, ascorbate peroxidase (APX, EC 1.11.1.11), oxidizes two AsA molecules to two MDA. APX belongs to class I haem peroxidases and is localized mainly in plastids (most isoforms), as well as in the cytoplasm and peroxisomes. A number of stromal APXs were also found in the mitochondrial matrix (Ishikawa, Shigeoka, 2008). MDA in the AGC can be reduced to AsA by monodehydroascorbate reductase (MDAR, EC 1.6.5.4), localized in the cytosol, peroxisomes, plastids and mitochondria. When reducing two molecules of MDA, MDAR oxidizes one molecule of

NADH or NADPH. DHA can be reduced non-enzymatically by glutathione, especially at the alkaline pH values found in the chloroplast stroma. Thioredoxins *f* and *m* also reduce DHA (Morell et al., 1997). In the AGC, dehydroascorbate reductase (DHAR, EC 1.8.5.1) is responsible for the reduction of DHA, while two molecules of reduced glutathione (GSH) are converted into an oxidized form – glutathione disulfide (GSSG). DHAR belongs to the glutathione S-transferase superfamily, although it is not capable of catalyzing glutathione conjugation or peroxide reduction. DHAR is localized in the cytosol, peroxisomes, plastids and mitochondria, and in a number of plants – in the apoplast and vacuoles (Ding et al., 2020). Glutathione reductase (GR, EC 1.8.1.7) catalyzes the reduction of glutathione at the expense of NADPH. GR functions in the cytoplasm, plastids and mitochondria, with 80 % of the activity registered in photosynthetic tissues occurring in chloroplasts (Gill et al., 2013). Data from proteomic analysis confirm the presence of GR in peroxisomes (Palma et al., 2009). GR is not only involved in the AGC, but also in the maintenance of sulfhydryl groups and the reduction of glutathione, oxidized directly by ROS, fatty acid peroxides and S-conjugates, etc. (Gill, Tuteja, 2010; Gill et al., 2013).

The AGC was formulated in relation to the antioxidant protection of the photosynthetic apparatus (Foyer, Halliwell, 1976). However, the cycle also operates in the cytosol and, partially, in peroxisomes and mitochondria. The AGC is important for plant adaptation to adverse environmental factors. Increased activity and gene expression of most AGC enzymes is characteristic for plants that are resistant to drought, salinity, non-optimal temperatures, heavy metals, phytopathogens, etc. Transgenic plants overexpressing genes of the AGC show increased resistance to stressors (Hasanuzzaman et al., 2019).

One of the common stress factors affecting plants is the deficiency (hypoxia) or complete absence (anoxia) of oxygen, occurring in wet or waterlogged soils, during inundation or flooding (Chirkova, Yemelyanov, 2018). ROS are formed in an oxygen-free environment, where they participate in the transduction of an anaerobic signal and trigger destructive processes (Blokhina et al., 2001). Subsequent oxidation of reduced products accumulated during anoxia leads to in-

creased generation of ROS and the development of post-anoxic oxidative damage (Blokhina et al., 2003; Shikov et al., 2020).

There are numerous data on the changes in the operation of AGC components under the influence of hypo- and anoxia. O₂ deficiency led to a decrease in the activity of all AGC enzymes in hypoxically grown rice and lotus seedlings (Ushimaru et al., 1992, 2001), and in wheat roots (Biemelt et al., 1998). However, no changes in enzymatic activity were detected in the roots of lupine (*Lupinus luteus*) (Garczarska, 2005). Prolonged waterlogging even stimulated most AGC enzymes in citrumelo (*Citrus paradisi* L. Macf. × *Poncirus trifoliata* L. Raf.) (Hossain et al., 2009). A decrease in AGC enzymes activity under flooding was observed in soybean roots – APX (Kausar et al., 2012), in Welsh onion – APX and GR (Yiu et al., 2009), and in cotton leaves – APX, DHAR and GR (Wang et al., 2019). Reaeration, on the contrary, stimulated most AGC enzymes (Ushimaru et al., 1992, 2001; Garczarska, 2005).

Only a handful of studies compared the operation of the AGC between plants with different levels of resistance to oxygen deficiency. In leguminous plants, pigeon pea (*Cajanus cajan*) and mung bean (*Vigna radiata*), varieties resistant to flooding were characterized by increased activity of APX and GR under hypoxia and reaeration, which was accompanied by upregulated expression of the corresponding genes (Sairam et al., 2009, 2011). An increase in APX activity during submergence was reported for tolerant slow-growing varieties of rice (Damanik et al., 2010) and ryegrass (*Lolium perenne*) (Liu, Jiang, 2015). Anoxia and reoxygenation caused greater ROS production and oxidative damage to lipids and proteins in seedlings of a non-resistant wheat plant (Chirkova et al., 1998; Shikov et al., 2022), and an increased activity of catalase and class III peroxidase in a resistant plant (rice) (Yemelyanov et al., 2022).

The objective of this study was to analyze the effects of anoxia and post-anoxic oxidative stress on the content of low-molecular-weight antioxidants and the activity of enzymes of the ascorbate-glutathione cycle in wheat and rice plants.

Materials and methods

Plant material. The objects of the study were 7-day-old wheat seedlings (*Triticum aestivum* L.) of the Leningradka variety and 10-day-old rice seedlings (*Oryza sativa* L.) of the Liman variety. Wheat caryopses were purchased from the Suida Breeding Station (Leningrad Region, Russia), rice seeds were provided by the Federal Rice Research Center (Krasnodar, Russia). Wheat was used as a plant that is not resistant to hypoxia, and rice was used as a hypoxia-tolerant one.

The seeds were surface-sterilized with a 5 % sodium hypochlorite solution, germinated, and seedlings were grown in hydroponic culture, as we previously described (Yemelyanov et al., 2020, 2022). Anaerobic conditions were created by passing nitrogen gas (oxygen content <0.01 %, Lentekhgaz, Russia) through chambers with plants, which were then hermetically sealed and placed in the dark to prevent the formation of oxygen in the light. Anaerobic conditions were checked using the Anaerotest[®] indicator (Merck, Germany). Exposure to a nitrogen atmosphere was 24 hours. Control

plants were placed in the dark under normoxic conditions. Next, to create post-anoxia, the experimental plants were removed from the anaerobic chambers and transferred into ambient air in the dark for 15 minutes, 1, 3 and 24 hours in experiments to determine low-molecular antioxidants, and 1 and 24 hours in the experiments to study the activity of enzymes of the ascorbate-glutathione cycle and expression of corresponding genes.

Concentrations of ascorbic acids and glutathione. Low-molecular-weight antioxidants of the AGC were extracted with chilled 5 % metaphosphoric acid from 1 g of shoots and roots after homogenization in liquid nitrogen (Blokhina et al., 2000). The extract was filtered and centrifuged for 30 minutes at 15,000 g. The content of ascorbate and dehydroascorbate was determined spectrophotometrically using the bipyridyl method (Knörzer et al., 1996). Reduced and oxidized glutathione was detected enzymatically (Law et al., 1983; Knörzer et al., 1996) using glutathione reductase (Sigma, USA).

Activities of AGC enzymes. To extract enzymes, shoots and roots of 10 seedlings were weighed and ground with a mortar and pestle in a chilled buffer (tissue : buffer ratio was 1:10, w/v) with the addition of quartz sand. All operations were performed at +4 °C. Determination of the protein content in the enzyme extract was carried out using the Bradford method (Bradford, 1976). When calculating enzyme activity, the autoxidation of substrates was taken into account, which was recorded without the addition of enzyme extract.

To extract *ascorbate peroxidase*, plant tissue was homogenized and extracted with 0.05 M K,Na-phosphate buffer (pH 7.0). After 15 minutes of centrifugation at 8,000 g, the supernatant was collected, and the pellet was resuspended with 1/2 of the original volume of buffer, then extracted for 20 minutes and centrifuged again. The combined fraction was used to determine enzyme activity by the decrease in absorption at 290 nm (spectrophotometer SP-26, LOMO, Russia) due to ascorbate oxidation (Nakano, Asada, 1981). The reaction medium for determining APX activity consisted of 40 µl of enzyme extract (7.0–23.2 µg of protein) in 0.05 M K,Na-phosphate buffer (pH 7.0), to which 0.5 ml of 5 mM ascorbate (Sigma) was added. The reaction was initiated by adding H₂O₂ (2.5 mM); distilled water was added to the control variant. The final volume of the reaction medium was 3 ml. Enzyme activity was calculated in micromoles of ascorbate oxidizable per 1 g of fresh weight per min using the extinction coefficient of 2.8 mM⁻¹ · cm⁻¹.

Intermediate AGC reductases (MDAR and DHAR) were extracted with 0.05 M K,Na-phosphate buffer (pH 7.3) containing 2 mM EDTA and 1 % cross-linked polyvinylpyrrolidone (Sigma). The homogenate was filtered and centrifuged at 15,000 g for 20 minutes.

MDAR activity was determined in a reaction coupled with ascorbate oxidase (Arrigoni et al., 1981). The composition of the reaction medium was 100 µl of 0.3 mM ascorbic acid, 200 µl of 0.3 mM NADH (Sigma), 0.5 units of ascorbate oxidase (Sigma), 300 µl of the sample (shoot extract) or 500 µl (root extract) and 0.05 M K,Na-phosphate buffer (pH 6.3). The final volume of the medium was 3 ml. The reaction was initiated by adding ascorbate oxidase. Optical density was measured at 340 nm with an SP-46 spectrophotometer

(LOMO, Russia). Enzyme activity was calculated in nanomoles of oxidizable NADH per 1 g of fresh weight per minute (NAD(P)H extinction coefficient was $6.22 \text{ mM}^{-1} \cdot \text{cm}^{-1}$).

DHAR activity was detected in the reaction medium containing 100 μl of 0.5 mM dehydroascorbic acid (Sigma), 100 μl of 1 mM reduced glutathione (Sigma), 50 μl of enzyme extract (from the shoots) or 100 μl (from the roots) and 0.05 M K,Na-phosphate buffer (pH 7.0) (Knörzer et al., 1996). The final volume was 3 ml. The reaction was initiated by adding DHA, which was dissolved in distilled water saturated with nitrogen gas immediately prior to measuring the activity. Optical density was measured at 265 nm with an SP-46 spectrophotometer. Enzyme activity was calculated in micromoles of reduced ascorbate per 1 g of fresh weight per minute.

Glutathione reductase was extracted with 0.1 M K,Na-phosphate buffer (pH 7.5) containing 2 mM EDTA and 1 % cross-linked polyvinylpyrrolidone. The homogenate was filtered and centrifuged at 15,000 g for 20 minutes. GR activity was determined in the following medium: 300 μl of extract, 100 μl of 0.5 mM oxidized glutathione (Sigma), 200 μl of 0.2 mM NADPH (Sigma) in 0.1 M K,Na-phosphate buffer (Rao et al., 1995). The final volume of the medium was 3 ml. The reaction was initiated by adding NADPH. Optical density was measured at 340 nm with an SP-46 spectrophotometer. Enzyme activity was calculated in nanomoles of NADPH oxidizable per 1 g of fresh weight per minute.

Gene expression was studied in rice seedlings, since the activity of most of the studied enzymes maintained or increased under stress. To design primers, we used the annotated rice genome databases (Rice Genome Annotation Project, <http://rice.uga.edu/>, last accessed 22 December 2023, and The rice annotation project database, <http://rapdb.dna.affrc.go.jp/tools/search/>, last accessed 22 December 2023), as well as the rice reference genome IRGSP-1.0 *Oryza sativa* var. *japonica* cv. Nipponbare, available at http://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001433935.1/ (last accessed 22 December 2023).

Nucleotide sequences of genes encoding all types of ascorbate peroxidases (8 genes), monodehydroascorbate reductases (5 genes), dehydroascorbate reductases (2 genes) and glutathione reductases (3 genes) were found. Coding DNA sequences (CDS) were examined. Among ascorbate peroxidases, several CDS were identified for the *OsAPX8* gene (2); among monodehydroascorbate reductases – for the *OsMDAR2* and *OsMDAR4* genes (2 each); among glutathione reductases – for *OsGR2* (3) and *OsGR3* (2). All CDS of the corresponding genes were aligned using the ClustalW algorithm in the MegAlign 5.05 program from the DNASTar suite. Primers were designed for the consensus regions closest to the 3' end of the CDS in the VectorNTI 8 program (Supplementary Materials 1 and 2)¹, i.e. the primers we developed allow to evaluate the expression of any alternatively spliced variants of the genes of interest. *OsTUB4*, encoding β -tubulin-4 and demonstrating the most stable expression, was used as a reference gene (Yemelyanov et al., 2022). The specificity of the primers was checked by searching for homology with the rice genome and transcriptome using the BLASTn algorithm

¹ Supplementary Materials 1–6 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Emel_Engl_28_1.pdf

on the NCBI database website (<http://blast.ncbi.nlm.nih.gov/>, last accessed 22 December 2023). Primers were ordered from the Beagle company (Russia, <http://www.biobeagle.com/>, last accessed 22 December 2023).

Methods for RNA isolation and purification, reverse transcription and quantitative real-time PCR (RT-PCR) have been described in details previously (Yemelyanov et al., 2022). RT-PCR was performed using kits with SYBRGreen dye (Syntol, Russia, <http://www.syntol.ru>, last accessed 22 December 2023) in a C1000 thermal cycler with a CFX96 optical module (Bio-Rad Laboratories, USA) according to the manufacturer's recommendations. We used the equipment of the Center for Molecular and Cell Technologies of Research park of St. Petersburg State University.

The $2^{-\Delta\text{Ct}}$ method was used to obtain the relative amount of transcripts from the difference in threshold amplification cycles (Ct) between the target and the reference gene (*OsTUB4*), and the $2^{-\Delta\Delta\text{Ct}}$ method was used to obtain the degree of change in the relative amount of transcripts of each gene (Livak, Schmittgen, 2001). Changes in the expression level were calculated relative to control (normoxic) values, taking them as one.

Statistical analysis. All experiments were carried out in 4–8 biological and 3 analytical replicates. Statistical data processing was performed using GraphPad Prism 8.0.1 for Windows. The graphs in figures show the average values and their standard errors. Values with different letters were significantly different at $p < 0.05$ (Tukey weighted mean). Heatmaps were generated using the tidyverse package (Wickham et al., 2019) in the R software environment (R Core Team, 2023). Asterisks on the heatmap indicate statistically significant differences from the control (Mann–Whitney U test, $p < 0.05$).

Results

The impact of anoxia and reoxygenation on the low-molecular-weight antioxidants of the AGC

The initial level of ascorbate in wheat was higher than in rice (twice as high in the shoots and 1.5 times in the roots, Fig. 1). The reduced form of ascorbate (AsA) dominated in the seedlings of both plants (80 and 70 % in wheat shoots and roots, respectively, and 60 % in the organs of the rice seedlings). The 24-hour anoxia caused a 4-fold decrease in the level of AsA in the shoots and a 5-fold decrease in the roots of the wheat seedlings (see Fig. 1, a, c), accompanied by the accumulation of dehydroascorbate (DHA) (increasing by 3.5 and 3 times, respectively). As a result, the oxidized form dominated, accounting for 80 % in the shoots and 90 % in the roots of wheat seedlings. In the shoots of rice, under the influence of anoxia, the content of AsA decreased by 10 %, and DHA remained unchanged (see Fig. 1, b). Meanwhile, in the roots, the level of both ascorbate forms increased by 40 % (see Fig. 1, d). A 15-minute reoxygenation led to a further depletion in the AsA content and the accumulation of DHA in wheat shoots. A more prolonged post-anoxic period (1–24 hours) led to the accumulation of AsA and a decrease in DHA levels (see Fig. 1, c), with both forms reaching control values. Notably, the total level of ascorbate (AsA + DHA) decreased from 2.3

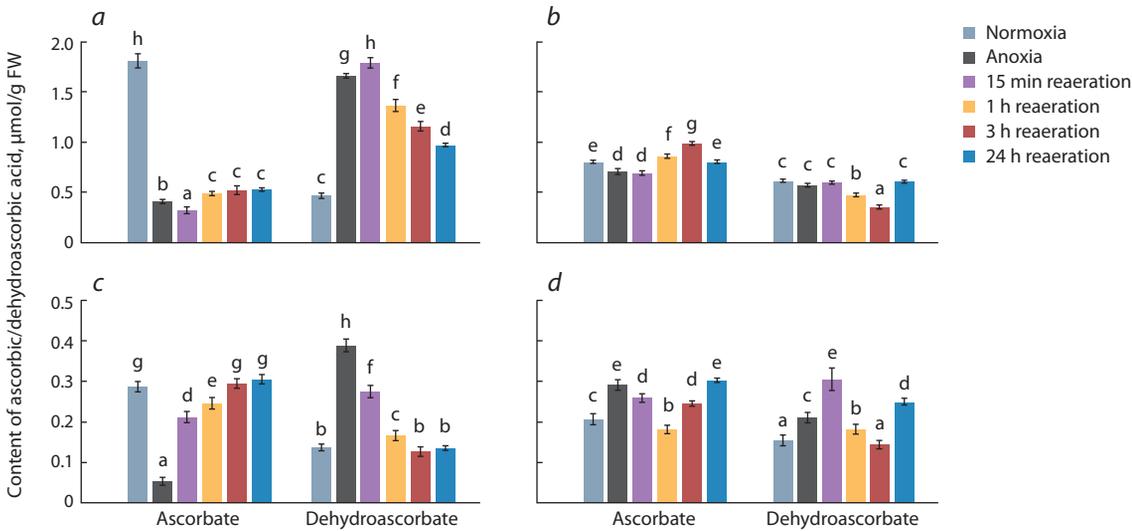


Fig. 1. Effect of 24 h anoxia and subsequent reoxygenation on the ascorbic acid (AsA) and dehydroascorbic acid (DHA) content in shoots (*a, b*) and roots (*c, d*) of wheat (*a, c*) and rice (*b, d*) seedlings.

Values with different letters (a–h) are significantly different at $p < 0.05$, according to Tukey's test.

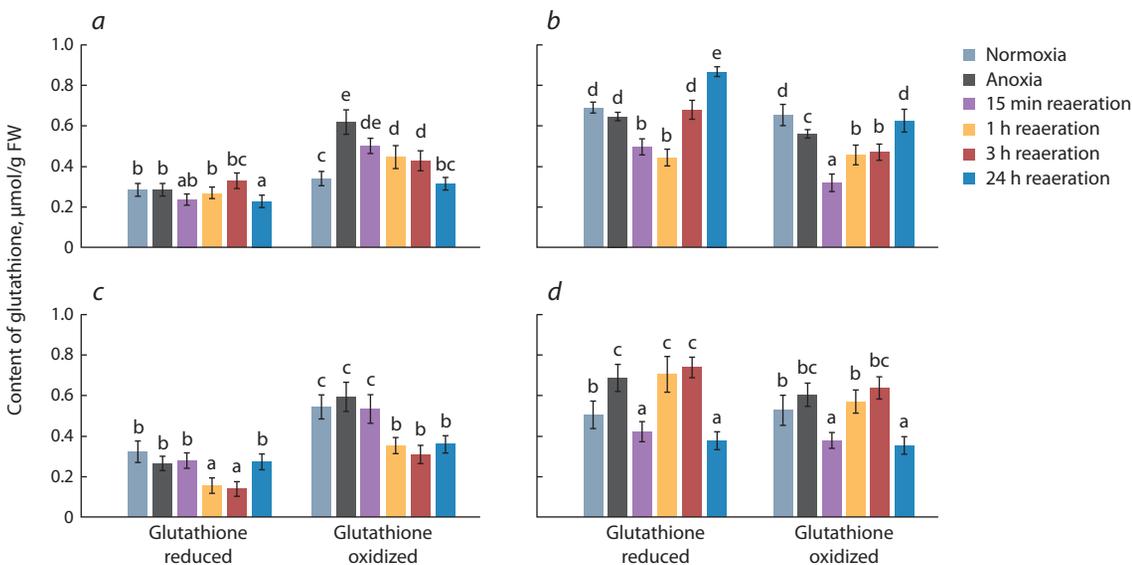


Fig. 2. Effect of 24 h anoxia and subsequent reoxygenation on the reduced (GSH) and oxidized (GSSG) glutathione content in shoots (*a, b*) and roots (*c, d*) of wheat (*a, c*) and rice (*b, d*) seedlings.

Values with different letters (a–e) are significantly different at $p < 0.05$, according to Tukey's test.

to 1.5 $\mu\text{mol/g}$ fresh weight in wheat shoots under the influence of anoxia and subsequent reoxygenation, and it was preserved in the roots. Changes of ascorbate in rice seedlings were in the opposite direction: in the shoots, the level of AsA/DHA initially increased/decreased (respectively) and then returned to the control level (see Fig. 1, *b*), while in the roots, conversely, the levels initially decreased/increased, but by the end of the experiment, both parameters exceeded control values (see Fig. 1, *d*). The total level of ascorbate in rice seedlings did not change, except for the 24-hour reoxygenation, when it increased 1.5 times in the roots.

Next, let us examine the dynamics of glutathione content. The initial concentrations of reduced glutathione (GSH) and

total glutathione (GSH+GSSG) were higher in rice seedlings, particularly in the shoots (Fig. 2). The content of both forms, GSH and GSSG was approximately the same. The ratio of GSH/GSSG was 1.3 in the shoots and 1.1 in the roots of rice, while in wheat, the oxidized form (GSSG) dominated, with a GSH/GSSG ratio of 0.6 and 0.9, respectively. The lack of oxygen did not affect the GSH level in wheat seedlings and rice shoots, whereas in rice roots, it increased by 25 % (see Fig. 2, *d*). The content of GSSG changed under anoxia only in the shoots of the studied plants: it increased twofold in wheat, while it decreased by 15 % in rice (see Fig. 2, *a, b*). The level of GSH decreased in wheat shoots after 24 hours of reoxygenation, and in the roots at shorter intervals (1–3 hours),

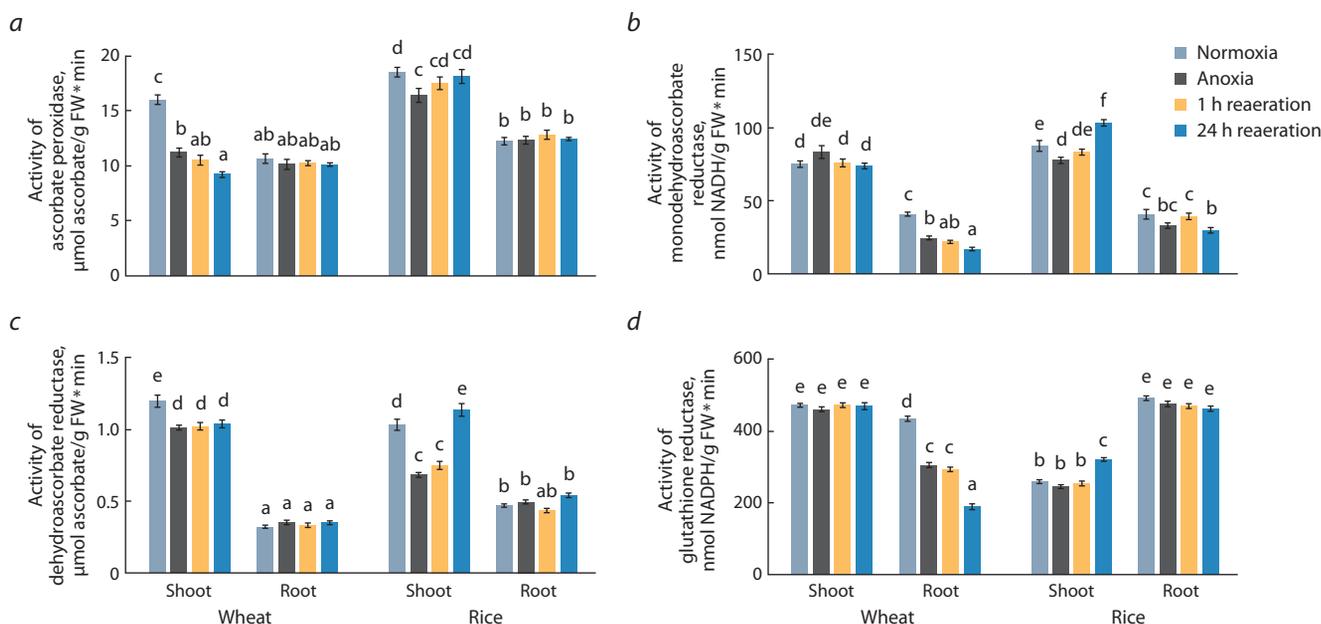


Fig. 3. Activity of enzymes of the ascorbate-glutathione cycle in the shoots and roots of wheat and rice seedlings under 24 h anoxia and subsequent reoxygenation: *a*, ascorbate peroxidase (APX); *b*, monodehydroascorbate reductase (MDAR); *c*, dehydroascorbate reductase (DHAR); *d*, glutathione reductase (GR).

Values with different letters (a–f) are significantly different at $p < 0.05$, according to Tukey's test.

after which it returned to the control level (see Fig. 2, *a*, *c*). During post-anoxia, the GSSG content in the shoots returned to the control level (normoxia) and was lower than under normoxic conditions in the roots. The level of total glutathione (GSH + GSSG) did not change significantly under stress, and the proportion of GSH decreased from 46 % (shoots) and 40 % (roots) under normoxia to 31–35 % under anoxia and short-term reoxygenation, after which it returned to control values. The post-anoxic period had a similar effect on the levels of both forms of glutathione in rice shoots. In the shoots, it decreased during short-term reoxygenation (15 minutes to 3 hours) and increased by 24 hours of post-anoxia (see Fig. 2, *b*). In rice roots, the content of both forms decreased after 15 minutes of reoxygenation, increased by 1–3 hours, and then decreased again by 24 hours of post-anoxia (see Fig. 2, *d*). The total glutathione level in rice decreased during anoxia and short-term reoxygenation (15 minutes), and then returned to the control normoxic level. The GSH/GSSG ratio in rice shoots was greater than one throughout all the experiment, indicating a predominance of GSH.

The impact of anoxia and reoxygenation on the enzymes of the AGC

Figure 3 depicts changes in the activity of high-molecular-weight components of the AGC. Interestingly, the baseline activity level of AGC enzymes, except for rice GR, was higher in the shoots of plants of both species (see Fig. 3). Anoxia and subsequent reoxygenation led to a significant downregulation of APX activity in wheat shoots, while in rice shoots, a 10 % decrease in activity under oxygen deficiency was followed by a return to the control level during post-anoxia (see Fig. 3, *a*). In the roots of both species, neither anoxia nor reoxygenation

caused changes in APX activity, similar to MDAR in wheat shoots (see Fig. 3, *b*). In wheat roots, the activity of MDAR decreased almost twofold as a result of oxygen deficiency and reoxygenation. In rice shoots, the change in MDAR activity was similar to that of APX, only after 24 hours of post-anoxia, the activity exceeded the control normoxic value. In rice roots, MDAR was consistently active throughout the entire experiment and was inhibited by 25 % only after 24 hours of reoxygenation (see Fig. 3, *b*).

The inactivation of DHAR was registered during anoxia in the shoots of both plants. However, in the case of rice during reoxygenation, the activity of the enzymes was restored and increased; however, this was not the case for wheat (see Fig. 3, *c*). In the roots of both plants, the activity of DHAR remained unchanged under all experimental conditions, similar to the activity of GR in wheat shoots and rice roots (see Fig. 3, *d*). In wheat roots, the enzyme was deactivated under stress conditions, while in rice, it was stimulated after 24 hours of reoxygenation (see Fig. 3, *d*).

The impact of anoxia and reoxygenation on the expression of genes encoding AGC enzymes in rice

An analysis of the expression of genes encoding AGC enzymes was conducted in rice tissues, where the activation of enzymes during post-anoxia was demonstrated. APX in rice is encoded by eight genes: *OsAPX1* and *2* encode cytosolic isoforms, *OsAPX3* and *4* – peroxisomal ones, *OsAPX5* and *6* – isoforms with dual plastid-mitochondrial localization, *OsAPX7* – the stromal isoform of plastids, *OsAPX8* – the thylakoid one (see Supplementary Material 1). In rice shoots, 24-hour anoxia resulted in an insignificant increase in the expression of peroxisomal *OsAPX3*, a 15-fold increase in the expression of

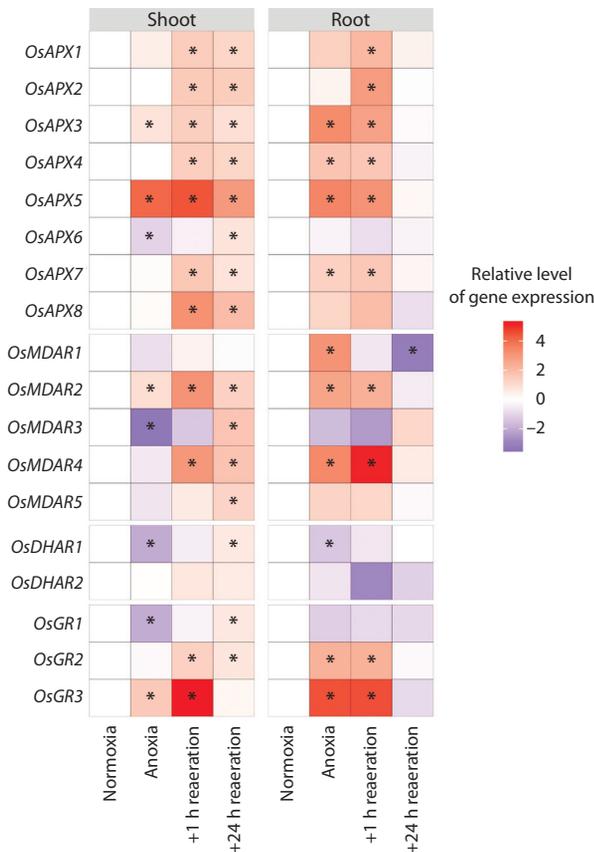


Fig. 4. Heatmaps of relative transcript levels of genes encoding enzymes of ascorbate-glutathione cycle in the shoots and roots of rice seedlings under 24 hours of anoxia and subsequent reoxygenation.

The expression level in the control (normoxia) is taken as one. Asterisks on the heatmap indicate statistically significant differences from the control (Mann-Whitney U test, $p < 0.05$).

OsAPX5, and a slight decrease in the expression of *OsAPX6*, the product of which is localized in plastids and mitochondria (Fig. 4, Supplementary Material 3).

Reoxygenation had the same impact on all genes of the *OsAPX* family, inducing the activation of their expression both after 1 and after 24 hours. Statistically significant differences in mRNA accumulation between an hour and a day of reoxygenation were detected for *OsAPX6*, the expression of which increased after a day of reoxygenation, as well as for *OsAPX7*, the expression of which also increased after an hour of reoxygenation but began to decrease after a day of reoxygenation. During post-anoxia, *OsAPX5* and *OsAPX8* were the most activated. In rice roots, the changes in expression were more pronounced (see Fig. 4, Supplementary Material 3). Due to anoxia action, the expression of peroxisomal forms (*OsAPX3* and *OsAPX4*), the plastid-mitochondrial form *OsAPX5*, and the stromal form *OsAPX7* increased. After an hour of reoxygenation, the expression of all genes, except for *OsAPX6*, increased; however, after a day of reoxygenation, the mRNA levels of all genes returned to their original values. The most active genes during reoxygenation in the roots were *OsAPX2*, *OsAPX3*, and *OsAPX5*. In both organs, the *OsAPX5* gene was the most responsive to anoxia and reoxygenation.

The family of genes encoding MDAR in rice consists of five genes. The products of the *OsMDAR1* and *OsMDAR2* genes are localized in peroxisomes, of *OsMDAR3* and *OsMDAR4*, in the cytosol, and of *OsMDAR5*, in plastids and mitochondria (see Supplementary Material 2). In the shoots, 24 hours of anoxia led to an increase in the expression of *OsMDAR2* and a decrease in the expression of *OsMDAR3* (see Fig. 4, Supplementary Material 4). After an hour of reoxygenation, the expression of *OsMDAR2* and *OsMDAR4* peaked, and after a day of reoxygenation, there was a tendency to return to the original level of expression. The accumulation of mRNA of *OsMDAR3* and *OsMDAR5* also increased after an hour of reoxygenation and continued to grow after 24 hours of reoxygenation. In the roots, anoxia activated the expression of *OsMDAR1*, *OsMDAR2*, and *OsMDAR4*. After an hour of reoxygenation, the quantity of *OsMDAR1* transcripts returned to the control normoxic level, while the expression of the *OsMDAR2* and *OsMDAR4* genes remained at an elevated anoxic level. After a day of reoxygenation, activity of all genes returned to the original level, and in the case of the *OsMDAR1* gene, it even dropped below it.

Between the two genes encoding DHAR in rice, changes were observed in the case of *OsDHAR1*. In both shoots and roots, expression followed a similar pattern: 24 hours of anoxia led to a twofold decrease in transcript levels; after an hour of reoxygenation, a tendency to return to control values emerged, which were successfully restored after a day of reoxygenation (see Fig. 4, Supplementary Material 5).

Three GR genes were identified in the rice genome: *OsGR1* and *OsGR3* encode isoforms of the enzyme with plastid-mitochondrial localization, while *OsGR2* encodes a cytosolic isoform (see Supplementary Material 2). The expression of *OsGR1* decreased, *OsGR2* remained unchanged, and *OsGR3* increased in the shoots after 24 hours of anoxia (see Fig. 4, Supplementary Material 6). After an hour of reoxygenation, the accumulation of *OsGR1* mRNA returned to the normoxic level, *OsGR2* increased, and *OsGR3* continued to rise, exceeding the control level by 30 times. After 24 hours of reoxygenation, the amount of *OsGR1* transcripts continued to increase, *OsGR2* remained at the same level, and *OsGR3* returned to the initial value. In the roots of rice, the changes in *OsGR2* and *OsGR3* were perfectly synchronized: an increase after anoxia, maintenance of the same level after an hour of reoxygenation, and restoration of the initial level after a day of reoxygenation. The expression of *OsGR1* remained unchanged. *OsGR3* was predominant in response to both anoxia and reoxygenation in both organs.

Discussion

The obtained results demonstrated that under normoxic conditions, the hypoxia-sensitive plant (wheat) accumulated higher levels of ascorbic acid, while the hypoxia-tolerant plant (rice) produced more glutathione (see Fig. 1, 2). After 24 hours of anoxia and short-term (15 minutes to 1 hour) reoxygenation, there was an accumulation of oxidized forms of ascorbate and glutathione in wheat tissues. The ratio of reduced forms to oxidized forms (AsA/DHA, GSH/GSSG) decreased below one, indicating the development of oxidative stress. It is important to note that in the roots of wheat, by 24 hours of post-anoxia,

the levels of the reduced antioxidants (AsA, GSH) returned to control levels (see Fig. 1, *c*, Fig. 2, *c*), whereas in the shoots, this did not occur. In wheat shoots, more DHA accumulated (see Fig. 1, *a*), and the total ascorbate level (AsA+DHA) decreased, indicating a greater development of oxidative processes. This is consistent with previously obtained data on higher oxidative damage to lipids in the shoots compared to the roots of wheat during anoxia and reoxygenation (Chirkova et al., 1998; Shikov et al., 2022).

In rice seedlings, the reduced forms of AGC antioxidants predominated both under normoxic conditions and during stress (see Fig. 1, *b, d*, Fig. 2, *b, d*). This suggests an effective antioxidant defence that mitigates the development of oxidative stress. Less oxidative damage to lipids and proteins (Chirkova et al., 1998; Shikov et al., 2022) and lower hydrogen peroxide (H₂O₂) production (Yemelyanov et al., 2022) were observed early in rice seedlings compared to wheat.

In studies performed by other authors, the accumulation of AsA and DHA, as well as a decrease in the levels of both forms of glutathione, were demonstrated in wheat under root anoxia (Biemelt et al., 1998). A short-term reoxygenation after anoxia caused a downregulation of the level of AsA, accumulation of DHA, and GSH. Though the total content remained unchanged, anoxia led to the depletion of the AsA pool and the accumulation of DHA, while reoxygenation stimulated the restoration of DHA in barley leaves (*Hordeum vulgare*) (Skutnik, Rychter, 2009), and in a suspension cell culture of *Arabidopsis thaliana* (Paradiso et al., 2016). During prolonged anoxia, the content of all forms of low-molecular-weight antioxidants of the AGC decreased in the roots of wheat and rice seedlings, as well as in the rhizomes of the *Iris* species, differing in resistance to hypoxia. However, in plants resistant to hypoxia, the depletion was less pronounced (Blokhina et al., 2000). The leaves of the moderately flood-tolerant citrume-lo variety CPB4475 contained high levels of AsA and GSH during prolonged waterlogging and subsequent reoxygenation (Hossain et al., 2009). Transgenic *A. thaliana* plants, tolerant to flooding and reoxygenation and overexpressing the *MYC2* gene, which encodes a transcription factor involved in jasmonic acid signalling, accumulated more AsA and GSH during hypoxia and reoxygenation compared to wild-type and *myc2* knockout mutants (Yuan et al., 2017). Therefore, it can be concluded that flood-tolerant plants are characterized by the prevalence of reduced forms of ascorbate and glutathione during oxygen deprivation and subsequent reoxygenation, which may result from their active reduction during the AGC operation.

The activity of AGC enzymes in the seedlings of the non-tolerant plant (wheat) under anoxia and reoxygenation either remained unchanged or was suppressed (APX and DHAR in the shoots, MDAR and GR in the roots) (see Fig. 3). As a result, the effective functioning of the AGC became impossible. The decrease in the activity of enzymes in the ascorbate part of the AGC (APX and DHAR) may be associated with the inability of wheat shoots to restore the pre-stress levels of AsA and DHA during oxygen deficiency and subsequent reoxygenation (see Fig. 1, *a*). At the same time, maintenance of the activity of these enzymes may contribute to achieving the pre-stress levels in the roots of wheat (see Fig. 1, *c*). In

the roots of the tolerant plant (rice), the activity of AGC enzymes under oxygen deficiency and post-anoxia also remained unchanged (except for MDAR at 24 hours of reoxygenation, see Fig. 3). In rice shoots, anoxia led to a decrease in the activity of APX, MDAR, and DHAR, and the reintroduction of oxygen stimulated all the enzymes of the cycle, resulting in the accumulation of reduced forms of AsA and GSH (see Fig. 1, 2).

Earlier studies have demonstrated that the deficiency or absence of O₂ resulted in a decrease in the activity of all AGC enzymes in wheat roots (Biemelt et al., 1998) and in the hypoxia-grown seedlings of rice and lotus (Ushimaru et al., 1992, 2001). There is also data on some of the enzymes of the AGC for several flood-intolerant species. Thus, oxygen deficiency suppressed the activity of APX, DHAR, and GR in the leaves of cotton (Wang et al., 2019) and barley (Skutnik, Rychter, 2009), APX and GR in Welsh onion roots (Yiu et al., 2009), and APX activity in soybean (Kausar et al., 2012). O₂ deficiency did not affect the activity of AGC enzymes in the roots of lupine (Garnczarska, 2005), while in the leaves of the moderately flood-tolerant citrus variety citrume-lo CPB4475, it stimulated the activity of most AGC enzymes, except MDAR (Hossain et al., 2009). In the suspension cell culture of *A. thaliana*, the activity of APX and GR was suppressed by anoxia and stimulated by reoxygenation, while the activities of MDAR and DHAR remained unchanged (Paradiso et al., 2016). GR was stimulated during post-hypoxia, while no stimulation of AGC enzymes was observed during post-anoxia in wheat roots (Biemelt et al., 1998). Reoxygenation stimulated the activity of all AGC enzymes in rice and lotus seedlings (Ushimaru et al., 1992, 2001), in lupine roots (Garnczarska, 2005), and GR in barley roots (Skutnik, Rychter, 2009). The activity of GR in barley shoots, as well as the activities of APX and DHAR in the whole plant, remained unchanged during post-anoxia (Skutnik, Rychter, 2009). A comparison of plants varying in flood tolerance revealed a higher activation of APX and GR during hypoxia and reoxygenation in the tolerant forms of pigeon pea, mung bean (Sairam et al., 2009, 2011), rice varieties (Damanik et al., 2010), and ryegrass (Liu, Jiang, 2015). It can be concluded that oxygen deficiency leads to the preservation or activation of AGC enzymes in flood-tolerant plants, and the reintroduction of O₂ into the environment stimulates their activity. This enables the reduction of oxidized forms of low-molecular-weight antioxidants of the cycle and ensures their efficient recyclization.

For a more in-depth analysis of the AGC operating in the anoxia-tolerant rice seedlings, we have investigated the expression of genes encoding the enzymes of the cycle. Although the activity of APX in rice shoots decreased during anoxia (see Fig. 3, *a*), the downregulation of expression was observed only in *OsAPX6*, encoding the plastid-mitochondrial isoform (see Fig. 4, Supplementary Material 3). Reversely, the activation of expression was observed for *OsAPX5*, encoding an enzyme of the same localization, and for peroxisomal *OsAPX3* in rice shoots in an anaerobic environment. The reoxygenation stimulated the accumulation of mRNA for all *OsAPX*, including the cytosolic *OsAPX1* and *OsAPX2*, which corresponded to the increase in APX activity to the control normoxic levels. The post-anoxic conditions tended to activate predominantly the

OsAPX5 and *OsAPX8* genes, which encode plastid isoforms. In the roots, enzyme activity remained unchanged, but the expression levels of peroxisomal forms (*OsAPX3* and *OsAPX4*) and plastid forms (*OsAPX5* and *OsAPX7*) increased already under anoxia, with the addition of the cytosolic (*OsAPX1* and *OsAPX2*) and the plastid form (*OsAPX8*) during post-anoxia. The greatest activation was observed for *OsAPX2*, *OsAPX3*, and *OsAPX5*, corresponding to the cytosolic, peroxisomal, and plastid-mitochondrial forms, respectively. The *OsAPX5* was activated more strongly than other genes and in both organs of the seedling.

There is limited information on the expression of *APX* genes in the literature. In soybean seedlings, flooding suppressed the expression of cytosolic *APX* genes, *GmAPX1* and *GmAPX2* (Nishizawa et al., 2013). In pigeon pea and mung bean, the hypoxia-tolerant forms were characterized by increased expression of genes encoding cytosolic *APX* under hypoxia, compared to the forms that are intolerant to flooding (Sairam et al., 2009, 2011). An increased expression of *OsAPX1* during flooding was detected in a submergence-tolerant rice variety carrying the *Sub1A* allele (Parlanti et al., 2011). The flood-tolerant transgenic line of *A. thaliana*, expressing the *MYC2* gene, and the wild-type plants were characterized by the flooding-induced activation of genes encoding AGC enzymes (*AtAPX2*, *AtMDHAR3*, *AtDHAR1*, and *AtGRI*) (Yuan et al., 2017). The mRNA level of *AtAPX2*, encoding the cytosolic isoform, also increased during anoxia in suspension cell culture of *Arabidopsis*, but it rose even more during a short-term reoxygenation, similar to other cytosolic forms (*AtAPX1* and *AtAPX6*). The other *APXs* genes were not investigated (Paradiso et al., 2016). We demonstrated the activation of *APX* genes encoding cytosolic forms in rice only under post-anoxia (see Fig. 4, Supplementary Material 3).

In our experiments, the activity of intermediate reductases of the AGC (MDAR and DHAR) changed similarly in the shoots: it decreased during anoxia and increased above the control level during reoxygenation. In the case of DHAR, this pattern coincided with changes in the expression of *OsDHAR1* (see Fig. 4, Supplementary Material 5). The decrease in MDAR activity in the shoots occurred alongside a decline in transcripts of *OsMDAR3*, encoding the cytosolic form, and the accumulation of transcripts of the peroxisomal form *OsMDAR2* (see Fig. 4, Supplementary Material 4). The expression of *OsMDAR2* and *OsMDAR4*, encoding the cytosolic isoforms, increased during a short-term reoxygenation. After 24 hours of post-anoxia, the expression of *OsMDAR3* and *OsMDAR5*, encoding the cytosolic and plastid-mitochondrial forms respectively, increased as well. In the roots, despite the absence of changes in the enzyme activity, anoxia activated the expression of *OsMDAR1*, *OsMDAR2*, and *OsMDAR4*. As in the shoots during reoxygenation, the quantity of mRNA of *OsMDAR2* and *OsMDAR4* increased. In the scientific literature, only the aforementioned article by L.-B. Yuan and co-authors (2017) was found, which investigated the expression of genes encoding AGC reductases during hypoxia and post-hypoxia. The maintenance of the GR activity during anoxia in both shoots and roots of rice, as well as the activation during reoxygenation in the shoots, is primarily associated with the expression of *OsGR2* and *OsGR3*, encoding the cytosolic

and plastid-mitochondrial isoforms, respectively (see Fig. 4, Supplementary Material 6).

Therefore, in most cases, oxygen deficiency induced the activity of the genes encoding the peroxisomal, plastid, and mitochondrial forms of AGC enzymes, and upon returning to the normoxic conditions, there was also an enhancement in the expression of the genes encoding the cytoplasmic forms. The involvement of different enzyme isoforms may be an important physiological mechanism for cell adaptation during the transition from oxygen-deficient conditions to reoxygenation. On the other hand, the addition of the cytoplasmic forms of enzymes to the organellar ones, which were already activated during anoxia, might be a consequence of increased oxidative stress upon the return of the oxygen levels to normoxic values during reoxygenation. In most cases, the activation of AGC enzymes during post-anoxia corresponded to the changes in their gene expression.

Conclusion

The comprehensive study of the AGC during anoxia and subsequent reoxygenation demonstrated the inactivation of *APX* and *DHAR* in the shoots, as well as *MDAR* and *GR* in the roots of the intolerant plant (wheat). This disruption compromised the efficient functioning of the cycle, leading to the accumulation of oxidized forms of ascorbate and glutathione and contributing to the development of significant oxidative damage. In the flood-tolerant plant (rice), the enzyme activity of the AGC was preserved under oxygen deficiency, and reoxygenation stimulated it, including the transcription level. This type of response possibly enables the recyclization of low-molecular-weight antioxidants in the cycle, providing antioxidant protection and preventing the development of oxidative stress. Thus, the resistance of plants to oxygen deficiency includes the resistance mechanisms to oxidative stress.

References

- Arrigoni O., Dipierro S., Borraccino G. Ascorbate free radical reductase, a key enzyme of the ascorbic acid system. *FEBS Lett.* 1981; 125(2):242-244. DOI 10.1016/0014-5793(81)80729-6
- Biemelt S., Keetman U., Albrecht G. Re-aeration following hypoxia or anoxia leads to activation of the antioxidative defense system in roots of wheat seedlings. *Plant Physiol.* 1998;116(2):651-658. DOI 10.1104/pp.116.2.651
- Blokhina O.B., Virolainen E., Fagerstedt K.V., Hoikkala A., Wähälä K., Chirkova T.V. Antioxidant status of anoxia-tolerant and -intolerant plant species under anoxia and reoxygenation. *Physiol. Plant.* 2000; 109(4):396-403. DOI 10.1034/j.1399-3054.2000.100405.x
- Blokhina O.B., Chirkova T.V., Fagerstedt K.V. Anoxic stress leads to hydrogen peroxide formation in plant cells. *J. Exp. Bot.* 2001; 52(359):1179-1190. DOI 10.1093/jxb/52.359.1179
- Blokhina O.B., Virolainen E., Fagerstedt K.V. Antioxidants, oxidative damage and oxygen deprivation stress: a review. *Ann. Bot.* 2003; 91(2):179-194. DOI 10.1093/aob/mcf118
- Bradford M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 1976;72(1-2):248-254. DOI 10.1016/0003-2697(76)90527-3
- Chirkova T., Yemelyanov V. The study of plant adaptation to oxygen deficiency in Saint Petersburg University. *Biol. Commun.* 2018; 63(1):17-31. DOI 10.21638/spbu03.2018.104
- Chirkova T.V., Novitskaya L.O., Blokhina O.B. Lipid peroxidation and antioxidant systems under anoxia in plants differing in their tolerance to oxygen deficiency. *Russ. J. Plant Physiol.* 1998;45(1):55-62

- Damanik R.I., Maziah M., Ismail M.R., Ahmad S., Zain A.M. Responses of the antioxidative enzymes in Malaysian rice (*Oryza sativa* L.) cultivars under submergence condition. *Acta Physiol. Plant.* 2010;32(4):739-747. DOI /10.1007/s11738-009-0456-3
- Ding H., Wang B., Han Y., Li S. The pivotal function of dehydroascorbate reductase in glutathione homeostasis in plants. *J. Exp. Bot.* 2020;71(12):3405-3416. DOI 10.1093/jxb/eraa107
- Foyer C.H., Halliwell B. The presence of glutathione and glutathione reductase in chloroplasts: a proposed role in ascorbic acid metabolism. *Planta.* 1976;133(1):21-25. DOI 10.1007/BF00386001
- Foyer C.H., Noctor G. Redox regulation in photosynthetic organisms: signaling, acclimation, and practical implications. *Antioxid. Redox Signal.* 2009;11(4):861-905. DOI 10.1089/ars.2008.2177
- Garczarska M. Response of the ascorbate-glutathione cycle to re-aeration following hypoxia in lupine roots. *Plant Physiol. Biochem.* 2005;43(6):583-590. DOI 10.1016/j.plaphy.2005.05.003
- Gill S.S., Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol. Biochem.* 2010;48:909-930. DOI 10.1016/j.plaphy.2010.08.016
- Gill S.S., Anjum N.A., Hasanuzzaman M., Gill R., Trivedi D.K., Ahmad I., Pereira E., Tuteja N. Glutathione and glutathione reductase: a boon in disguise for plant abiotic stress defense operations. *Plant Physiol. Biochem.* 2013;70:204-212. DOI 10.1016/j.plaphy.2013.05.032
- Halliwell B. Reactive species and antioxidants. Redox biology is a fundamental theme of aerobic life. *Plant Physiol.* 2006;141(2):312-322. DOI 10.1104/pp.106.077073
- Hasanuzzaman M., Nahar K., Anee T.I., Fujita M. Glutathione in plants: biosynthesis and physiological role in environmental stress tolerance. *Physiol. Mol. Biol. Plants.* 2017;23(2):249-268. DOI 10.1007/s12298-017-0422-2
- Hasanuzzaman M., Bhuyan M.H.M.B., Anee T.I., Parvin K., Nahar K., Mahmud J.A., Fujita M. Regulation of ascorbate-glutathione pathway in mitigating oxidative damage in plants under abiotic stress. *Antioxidants.* 2019;8(9):384. DOI 10.3390/antiox8090384
- Hossain Z., Lopez-Climent M.F., Arbona V., Perez-Clemente R.M., Gomez-Cadenas A. Modulation of the antioxidant system in citrus under waterlogging and subsequent drainage. *J. Plant Physiol.* 2009;166(13):1391-1404. DOI 10.1016/j.jplph.2009.02.012
- Ishikawa T., Shigeoka S. Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. *Biosci. Biotechnol. Biochem.* 2008;72(5):1143-1154. DOI 10.1271/bbb.80062
- Kausar R., Hossain Z., Makino T., Komatsu S. Characterization of ascorbate peroxidase in soybean under flooding and drought stresses. *Mol. Biol. Rep.* 2012;39(12):10573-10579. DOI 10.1007/s11033-012-1945-9
- Knörzer O.C., Durner J., Boger P. Alterations in the antioxidative system of suspension cultured soybean cells (*Glycine max*) induced by oxidative stress. *Physiol. Plant.* 1996;97(2):388-396. DOI 10.1034/j.1399-3054.1996.970225.x
- Law M.Y., Charles S.A., Halliwell B. Glutathione and ascorbic acid in spinach (*Spinacia oleracea*) chloroplasts. The effect of hydrogen peroxide and of paraquat. *Biochem. J.* 1983;210(3):899-903. DOI 10.1042/bj2100899
- Liu M., Jiang Y. Genotypic variation in growth and metabolic responses of perennial ryegrass exposed to short-term waterlogging and submergence stress. *Plant Physiol. Biochem.* 2015;95:57-64. DOI 10.1016/j.plaphy.2015.07.008
- Livak K.J., Schmittgen T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods.* 2001;25(4):402-408. DOI 10.1006/meth.2001.1262
- Morell S., Follmann H., De Tullio M., Haberlein I. Dehydroascorbate and dehydroascorbate reductase are phantom indicators of oxidative stress in plants. *FEBS Lett.* 1997;414(3):567-570. DOI 10.1016/S0014-5793(97)01074-0
- Nakano Y., Asada K. Hydrogen peroxide is scavenged by ascorbate-specific peroxidase in spinach chloroplasts. *Plant Cell Physiol.* 1981;22(5):867-880. DOI 10.1093/oxfordjournals.pcp.a076232
- Nishizawa K., Hiraga S., Yasue H., Chiba M., Tougou M., Nanjo Y., Komatsu S. The synthesis of cytosolic ascorbate peroxidases in germinating seeds and seedlings of soybean and their behavior under flooding stress. *Biosci. Biotechnol. Biochem.* 2013;77(11):2205-2209. DOI 10.1271/bbb.130384
- Noctor G., Foyer C.H. Ascorbate and glutathione: keeping active oxygen under control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 1998;49:249-279. DOI 10.1146/annurev.arplant.49.1.249
- Palma J.M., Corpas F.J., Del Rio L.A. Proteome of plant peroxisomes: new perspectives on the role of these organelles in cell biology. *Proteomics.* 2009;9(9):2301-2312. DOI 10.1002/pmic.200700732
- Paradiso A., Caretto S., Leone A., Bove A., Nisi R., De Gara L. ROS production and scavenging under anoxia and re-oxygenation in *Arabidopsis* cells: a balance between redox signaling and impairment. *Front. Plant Sci.* 2016;7:1803. DOI 10.3389/fpls.2016.01803
- Parlanti S., Kudahettige N.P., Lombardi L., Mensuali-Sodi A., Alpi A., Perata P., Pucciariello C. Distinct mechanisms for aerenchyma formation in leaf sheaths of rice genotypes displaying a quiescence or escape strategy for flooding tolerance. *Ann. Bot.* 2011;107(8):1335-1343. DOI 10.1093/aob/mcr086
- R Core Team. R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. 2023. (<https://www.R-project.org/>)
- Rao M.V., Hale B.A., Ormrod D.P. Amelioration of ozone-induced oxidative damage in wheat plants grown under high carbon dioxide. Role of antioxidant enzymes. *Plant Physiol.* 1995;109(2):421-432. DOI 10.1104/pp.109.2.421
- Sairam R.K., Kumutha D., Ezhilmathi K., Chinnusamy V., Meena R.C. Waterlogging induced oxidative stress and antioxidant enzyme activities in pigeon pea. *Biol. Plant.* 2009;53(3):493-504. DOI 10.1007/s10535-009-0090-3
- Sairam R.K., Dharmar K., Lekshmy S., Chinnusamy V. Expression of antioxidant defense genes in mung bean (*Vigna radiata* L.) roots under water-logging is associated with hypoxia tolerance. *Acta Physiol. Plant.* 2011;3(3):735-744. DOI 10.1007/s11738-010-0598-3
- Shikov A.E., Chirkova T.V., Yemelyanov V.V. Post-anoxia in plants: consequences, and possible mechanisms. *Russ. J. Plant Physiol.* 2020;67(1):45-59. DOI 10.1134/S1021443720010203
- Shikov A.E., Chirkova T.V., Yemelyanov V.V. Functions of reactive oxygen species in plant cells under normal conditions and during adaptation. *Ecol. Genet.* 2021;19(4):343-363. DOI 10.17816/ecogen75975
- Shikov A.E., Lastochkin V.V., Chirkova T.V., Mukhina Z.M., Yemelyanov V.V. Post-anoxic oxidative injury is more severe than oxidative stress induced by chemical agents in wheat and rice plants. *Acta Physiol. Plant.* 2022;44(9):90. DOI 10.1007/s11738-022-03429-z
- Skutnik M., Rychter A.M. Differential response of antioxidant systems in leaves and roots of barley subjected to anoxia and post-anoxia. *J. Plant Physiol.* 2009;166(9):926-937. DOI 10.1016/j.jplph.2008.11.010
- Ushimaru T., Shibasaki M., Tsuji H. Development of O₂ detoxification system during adaptation to air of submerged rice seedlings. *Plant Cell Physiol.* 1992;33(8):1065-1071. DOI 10.1093/oxfordjournals.pcp.a078357
- Ushimaru T., Kanematsu S., Katayama M., Tsujid H. Antioxidative enzymes in seedlings of *Nelumbo nucifera* germinated under water. *Physiol. Plant.* 2001;112(1):39-46. DOI 10.1034/j.1399-3054.2001.1120106.x
- Wang H., Chen Y., Hu W., Snider J.L., Zhou Z. Short-term soil-waterlogging contributes to cotton cross tolerance to chronic elevated temperature by regulating ROS metabolism in the subtending leaf. *Plant Physiol. Biochem.* 2019;139:333-341. DOI 10.1016/j.plaphy.2019.03.038

- Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., François R., Grolemond G., Hayes A., Henry L., Hester J., Kuhn M., Pedersen T.L., Miller E., Bache S.M., Müller K., Ooms J., Robinson D., Seidel D.P., Spinu V., Takahashi K., Vaughan D., Wilke C., Woo K., Yutani H. Welcome to the tidyverse. *J. Open Source Softw.* 2019;4(43):1686. DOI 10.21105/joss.01686
- Yemelyanov V.V., Lastochkin V.V., Chirkova T.V., Lindberg S.M., Shishova M.F. Indoleacetic acid levels in wheat and rice seedlings under oxygen deficiency and subsequent reoxygenation. *Biomolecules.* 2020;10(2):276. DOI 10.3390/biom10020276
- Yemelyanov V.V., Lastochkin V.V., Prikaziuk E.G., Chirkova T.V. Activities of catalase and peroxidase in wheat and rice plants under conditions of anoxia and post-anoxic aeration. *Russ. J. Plant Physiol.* 2022;69(6):1-15. DOI 10.1134/S1021443722060036
- Yiu J.-C., Liu C.-W., Fang D.Y., Lai Y.-S. Waterlogging tolerance of Welsh onion (*Allium fistulosum* L.) enhanced by exogenous spermidine and spermine. *Plant Physiol. Biochem.* 2009;47(8):710-716. DOI 10.1016/j.plaphy.2009.03.007
- Yuan L.-B., Dai Y.-S., Xie L.-J., Yu L.-J., Zhou Y., Lai Y.-X., Yang Y.-C., Xu L., Chen Q.-F., Xiao S. Jasmonate regulates plant responses to postsubmergence reoxygenation through transcriptional activation of antioxidant synthesis. *Plant Physiol.* 2017;173(3):1864-1880. DOI 10.1104/pp.16.01803

ORCID

V.V. Yemelyanov orcid.org/0000-0003-2323-5235
E.G. Prikaziuk orcid.org/0000-0002-7331-7004
V.V. Lastochkin orcid.org/0000-0001-9540-497X
O.M. Aresheva orcid.org/0009-0000-6024-2966
T.V. Chirkova orcid.org/0000-0002-2315-0816

Acknowledgements. The research was funded by the Russian Science Foundation, grant number 22-24-00484, <https://rscf.ru/en/project/22-24-00484/>. The investigation was performed using equipment of the Research Park "Center for Molecular and Cell Technologies" at Saint Petersburg State University, and the paper is dedicated to the 300th anniversary of St. Petersburg State University.

Conflict of interest. The authors declare no conflict of interest.

Received August 18, 2023. Revised November 27, 2023. Accepted November 28, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Morphological and molecular analysis of rose cultivars from the Grandiflora and Kordesii garden groups

S.S. Yudanova , O.V. Dorogina, O.Yu. Vasilyeva

Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
 judanowa.sophia@yandex.ru

Abstract. The breeding of remontant rose cultivars that are resistant to diseases and adverse conditions, with high decorative value and continuous flowering is the most important task during work with the gene pool of garden roses. Currently, intercultivar hybridization within a single garden group has largely outlived its usefulness. It is necessary to breed for highly decorative forms or cultivars that have outstanding resistance, morphological characters and patterns of seasonal rhythms, and use these plants as parental forms in further breeding. This study represents a comparative analysis of rose cultivars from two garden groups, Grandiflora (Gurzuf, Lezginka, Korallovy Syurpriz, Queen Elizabeth, Komsomolsky Ogonyok, Love) and Rosa Kordesii (Letniye Zvyozdy, Dortmund, Gutsulochka). These cultivars proved themselves during many years of testing in harsh climatic conditions. The objectives of the study were to determine the genetic relationship within the groups and to assign phenotypically different cultivars to one or another garden group. The analysis was carried out by morphological, phenological and ISSR markers. According to the phenological observations on the Grandiflora cultivars, Komsomolsky Ogonyok had later budding and flowering stages. Polymorphic data generated from the ISSR markers showed that this cultivar was the most distant from the others and formed a separate cluster on the dendrogram. A comparison of the morphological characters (flower diameter, number of petals, peduncle length, bush height) showed a significant difference ($p < 0.05$) between Komsomolsky Ogonyok and the other Grandiflora cultivars. A dendrogram based on a molecular analysis showed a lack of close relationships between Komsomolsky Ogonyok and the Kordesii group, which formed a separate cluster. A pairwise comparison of the morphological characters in Komsomolsky Ogonyok with the Kordesii group revealed a significant ($p < 0.05$) difference in three of the four characters studied. The exceptions were flower diameter when comparing with Dortmund and Letniye Zvyozdy and peduncle length when comparing with Gutsulochka. Although Komsomolsky Ogonyok has a pattern of seasonal development similar to Dortmund in the Kordesii group, the molecular analysis did not assign the former to this group of roses. The cultivars that have valuable characters that no average rose does and that are phenotypically different from such roses represent the most valuable breeding material.

Key words: *Rosa* L.; grandiflora; Rosa Kordesii; ISSR markers; morphological characters; phenological observations.

For citation: Yudanova S.S., Dorogina O.V., Vasilyeva O.Yu. Morphological and molecular analysis of rose cultivars from the Grandiflora and Kordesii garden groups. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(1):55-62. DOI 10.18699/vjgb-24-07

Морфологический и молекулярный анализ сортов роз из садовых групп грандифлора и розы Кордеса

С.С. Юданова , О.В. Дорогина, О.Ю. Васильева

Центральный сибирский ботанический сад Сибирского отделения Российской академии наук, Новосибирск, Россия
 judanowa.sophia@yandex.ru

Аннотация. Поиск зимостойких, устойчивых к грибным болезням сортов, характеризующихся высокой декоративностью, ремонтантным и продолжительным цветением, является важнейшей задачей при работе с коллекционным генофондом садовых роз. В настоящее время межсортовая гибридизация роз в пределах одной садовой группы во многом исчерпала себя. Требуется поиск высокодекоративных форм или сортов, выделяющихся по резистентности, морфологическим и ритмологическим признакам для использования в селекции в качестве родительских форм. В работе выполнен сравнительный анализ сортов из двух садовых групп – грандифлора (Гурзуф, Лезгинка, Коралловый сюрприз, Queen Elizabeth, Комсомольский огонёк, Love) и кордесии (Летние звёзды, Dortmund, Гуцулочка). Эти сорта хорошо показали себя в течение многих лет испытаний в суровых климатических условиях. Целью исследования было определение степени родства внутри групп и установление возможной принадлежности фенотипически различающихся сортов к одной из групп. Анализ проводили по морфологическим, фенологическим признакам, а также с помощью ISSR-маркеров. По результатам фенологических наблюдений в группе грандифлора выделился сорт Комсомольский огонёк: более позднее вступление

в фазы бутонизации и цветения. Данные полиморфизма, полученные на основании ISSR-маркирования, показали, что он удален от других сортов, образуя на дендрограмме отдельный кластер. Сравнение сортов роз по морфологическим признакам (диаметр цветка, количество лепестков, длина цветоноса, высота куста) также свидетельствует о достоверных отличиях ($p < 0.05$) сорта Комсомольский огонёк от остальных сортов группы грандифлора. Дендрограмма, построенная на основании молекулярного анализа, показала отсутствие близкого родства сорта Комсомольский огонёк и группы кордезии, которые формировали отдельный кластер. При попарном сравнении морфологических показателей сорта Комсомольский огонёк и группы кордезии обнаружено достоверное различие ($p < 0.05$) по трем из четырех изученных признаков, за исключением диаметра цветка (при сравнении с сортами Dortmund и Летние звёзды) и длине цветоноса (при сравнении с Гуцулочкой). Несмотря на то что Комсомольский огонёк по феноритмике схож с сортом Dortmund из группы кордезии, молекулярный анализ не позволяет отнести его к данной группе роз. Такие сорта, фенотипически отличающиеся от общей массы и обладающие рядом ценных признаков, являются ценнейшим селекционным материалом. Ключевые слова: *Rosa* L.; грандифлора; розы Кордеса; ISSR-маркеры; морфологические признаки; фенологические наблюдения.

Introduction

Roses (*Rosa* L.) are among the oldest plants cultivated by man not only for decorative use but also for perfumery, medical and culinary purposes. The genus includes about 200 species; however, only 10–15 of them have contributed to the garden groups of modern roses (Cairns, 2007). According to the modern classification, the world's entire collection represented by 40,000 cultivars is subdivided into 36 horticultural groups (Annotated Catalog ..., 2018; Plugatar et al., 2019). The most valuable quality of the cultivars from the Tea-Hybrid, Floribunda, Grandiflora, *Rosa* Kordesii, Polyanthus and Miniature groups is their ability for remontant flowering (Klimenko, 2010; Gorodnyaya, 2014; Tyshchenko, 2015), which is biologically conditioned by the presence in their hereditary basis of the genetic material from the evergreen species of section *Indicae* that do not tolerate winter well and are not prone to winter dormancy.

One of the main criteria for selecting garden rose cultivars promising for cultivation in harsh climatic conditions is flowering of their annual shoots (Vasilyeva, 1999). However, this biological feature is not characteristic of all garden groups and not of all species of the same garden group, e. g. most cultivars of large-flowered climbing (LCI) roses produce generative shoots on perennial shoot formation systems (SFSs) that die almost every year due to severe winters (Pashina, 2011; Kapelyan, 2017; Plugatar et al., 2018). For that reason, in terms of rose gardens grown in continental climate, along with the Tea-Hybrid and Floribunda, such groups as Grandiflora and *Rosa* Kordesii are of great interest.

The Grandiflora (Gr.) roses were bred in the 1950s solely based on their morphological characters and with no regard to the origin, they resulted from crossed Floribunda and Tea-Hybrid roses. They are praised for their abundant and remontant flowering, as that of the Floribunda, as well as for the long straight shoots with large flowers of different colors resembling those of the Tea-Hybrid. Unlike the latter, the Grandiflora roses grow not single flowers but small-flower inflorescences. The most important character of the group has been their growth strength and higher winter hardiness if compared to Tea-Hybrid roses.

Kordes Roses, or *Rosa* Kordesii, is a relatively young garden group, selected from a spontaneous *Rosa rugosa* × *Rosa wichuraiana* hybrid by the W. Kordes' Sohne Company,

whose breeding priority has been selection of unpretentious and winter-hardy forms. Their features are abundant flowering from June to late fall, high winter hardiness and increased resistance to diseases (Bardakova, 2017; Adritskaya, Kapelyan, 2022). In harsh climates, these roses grow flowers on annual shoots, which can make them a substitute for the climbing roses flowering on perennial SFSs and poorly surviving Siberian winters.

A striking representative of *Rosa* Kordesii is the Dortmund roses that are often used in breeding as a parental form for being resistant to fungal diseases. In the Nikitsky Botanical Garden, Z.K. Klimenko cultivated such *Rosa* Kordesii cultivars as Letniye Zvyozdy and Gutsulochka (Klimenko, Rubtsova, 1986) that have proved to be highly decorative and stable in harsh climatic conditions. Queen Elizabeth is the most popular representative of the Grandiflora group, known for its decorative features and complex resistance, and for these reasons it has been repeatedly used in breeding. Among the Russian members of the group, Komsomolsky Ogonyok (Charlotte Wheatcroft × Gloria Dei cross) is the most popular. Its long-term trials carried out in the Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences (CSBG SB RAS) demonstrated that it was phenotypically different from the Grandiflora group but had similarities with *Rosa* Kordesii. For that reason, the research presented in this paper was to evaluate the cultivars from the Grandiflora and *Rosa* Kordesii groups based on their morphological and molecular genetic characters in order to determine the kinship within the groups and to possibly establish whether the phenotypically distinguished cultivars belonged to one of the groups mentioned.

An additional goal was investigating CSBG SB RAS collection's gene pool for valuable forms to perform further breeding. Considering Siberia's harsh continental climate, we searched for cultivars of high winter hardiness, resistant to fungal diseases and characterized by remontant and prolonged flowering.

To evaluate the collection's genetic polymorphism, ISSR (inter simple sequence repeats) analysis was employed. The technique interprets the DNA sequences flanked by micro-satellite loci, has good reproducibility and does not require cloning and sequencing of DNA fragments for primer selection, thereby significantly reducing its cost and labor intensity.

Considering that the number of microsatellite repeats is very high in the genome in both animals and plants, this method is a very convenient tool for genetic analysis (Amom, Nongdam, 2017; Dorogina, Zhmud, 2020).

Materials and methods

In our study, we investigated the roses of the Grandiflora (Gurzuf, Lezginka, Korollovy Syurpriz, Queen Elizabeth, Komsomolsky Ogonyok, Love) and Rosa Kordesii (Letniye Zvyozdy, Dortmund, Gutsulochka) groups. The Grandiflora roses had (1) small inflorescences of large flowers in different colors; (2) tall bushes reaching up to 2 m in height if grown in southern Russia; (3) large and glossy leaves; (4) abundant, remontant flowering; (5) sufficiently high winter hardiness, which is favorable for Siberia. The Kordesii roses, on the other hand, were characterized by high winter hardiness, resistance to diseases and abundant long flowering. In a continental climate, this group can partially substitute climbing roses because they flower on annual shoots.

The roses' morphobiological characters such as flower diameter, number of petals, peduncle length, and bush height were studied during five summer seasons from 2017 to 2021. The study was carried out at CSBG SB RAS's Collections of Living Plants in Open and Protected Grounds, USU 440534 (54°49'13.8" N 83°06'13.3" E). To evaluate the morphobiological characters, standard methods were applied (Methodology of State Variety Testing..., 1968; Klimentko et al., 2019; Suprun, 2021). Phenological observations were carried out according to I.N. Beideman's method (1974) with modifications (Fomina, 2012). Student's *t*-test was used to confirm the reliability of the differences obtained for metric characters (Haynes, 2013). Average mean and standard error $M \pm \bar{x}$ were calculated using 20 plants.

For DNA isolation, the CTAB method with some modifications was used (Doyle J.J., Doyle J.L., 1987). Amplification was carried out according to the following program: primary denaturation for 2 min at 95 °C; 35 amplification cycles – denaturation for 20 sec at 94 °C, primer annealing for 45 sec, elongation for 1.5 min at 72 °C; final elongation for 7 min at 72 °C. PCR and further electrophoretic separation of amplification products were performed in 1–1.5 % agarose gel in 1×TBE buffer according to standard methods (Vasilyeva et al., 2020). The list of ISSR primers used in this work, their characteristics and annealing temperatures are given in Table 1.

Quantitative assessment of marker polymorphism and determination of the level of divergence between the studied forms were performed in a binary matrix where the presence or absence of PCR fragments of equal size was denoted as 1 or 0. For statistical data processing, the TREECON software (Van de Peer, Wachter, 1994) was used. The genetic distances were calculated as:

$$GD_{xy} = 1 - 2N_{xy}/(N_x + N_y).$$

Here, N_{xy} is the number of total fragments for samples x and y , N_x and N_y are the number of fragments for samples x and y , respectively (Nei, Li, 1979).

The nearest neighbors algorithm with bootstrap support of at least 100 was employed to build ISSR marker distribution dendrograms. The polymorphism level (P, %) of each primer was calculated by the formula:

$$P = 100 \times N_p/N,$$

where N_p is the number of polymorphic fragments and N is the total number of fragments.

Results

The vegetation period of roses in Siberia includes the following stages: aftergrowth; current-year shooting; budding; first (I) and second (II) flowering; defloration. As for more favorable climatic conditions in some, mainly southern, regions of Russia, garden roses have a third (III) flowering stage. The long-time average annual data accumulated during phenological observations of 2017–2022 demonstrated that the Grandiflora roses had earlier shooting regrowth than the Kordesii ones, despite the winter shelter being removed from the entire collection at the same time (Table 2). The time required for the shoots to produce the first flowers is more extended in the Kordesii roses, which is due to the more powerful shoots with a greater number of internodes than those of the Grandiflora; however, reflowering was less prolonged in the Kordesii roses.

Among the studied cultivars, the earliest flowering was observed in the foreign Grandiflora roses (Love and Queen Elizabeth). The time Queen Elizabeth entered the flowering phase on average came at the 61st day after removing winter shelters and beginning of vegetation, which was significantly ($p < 0.01$) different from all the cultivars studied, except for Love: its time to flowering was close to that of Queen Elizabeth. The other cultivars bloomed in 72–79 days from the

Table 1. Characteristics of ISSR primers used to study Grandiflora and Kordes Roses

Primer	Sequence, 5'–3'	Annealing temperatures, °C
HB12	(CAC) ₃ GC	42
17899B	(CA) ₆ GG	42
UBC807	(AG) ₈ T	52
UBC834	(AG) ₈ YT	60
UBC855	(AC) ₈ YT	50
M2	(AC) ₈ YG	50

Table 2. Long-term phenological data of rose cultivars from the Grandiflora and Kordes Roses groups (Novosibirsk, 2017–2021)

Cultivar	Spring sprouting	Bud formation	First bloom cycle			Second bloom cycle		
			Beginning	End	Duration, days	Beginning	End	Duration, days
Grandiflora group								
Gurzuf	07.05 ± 2	17.06 ± 2	09.07 ± 2	24.07 ± 4	16	12.08 ± 4	07.10 ± 2	57
Lezginka	04.05 ± 4	15.06 ± 2	12.07 ± 2	30.07 ± 3	19	17.08 ± 2	22.09 ± 3	37
Korallovy Syurpriz	05.05 ± 3	09.06 ± 2	07.07 ± 3	27.07 ± 2	21	12.08 ± 3	10.10 ± 2	60
Queen Elizabeth	02.05 ± 3	05.06 ± 3	26.06 ± 3	14.07 ± 2	19	03.08 ± 2	02.10 ± 3	61
Komsomolsky Ogonyok	04.05 ± 2	18.06 ± 2	09.07 ± 2	25.07 ± 3	17	21.08 ± 2	28.09 ± 3	39
Love	05.05 ± 3	07.06 ± 3	29.06 ± 4	22.07 ± 2	24	14.08 ± 3	03.10 ± 4	51
Rosa Kordesii group								
Letniye Zvyozdy	10.05 ± 3	20.06 ± 2	09.07 ± 3	28.07 ± 2	20	18.08 ± 3	21.09 ± 3	35
Dortmund	08.05 ± 2	23.06 ± 4	14.07 ± 2	31.07 ± 3	18	24.08 ± 4	26.09 ± 3	33
Gutsulochka	10.05 ± 4	20.06 ± 3	08.07 ± 3	03.08 ± 3	27	19.08 ± 3	19.09 ± 3	32

Table 3. Morphological characters of rose cultivars from the Grandiflora and Kordes Roses groups

Cultivar	Flower diameter, cm	Number of petals per flower	Peduncle length, cm	Bush height, cm
Grandiflora group				
Queen Elizabeth	8.75 ± 0.22	29.92 ± 0.71	41.33 ± 2.20	63.67 ± 1.36
Korallovy Syurpriz	9.27 ± 0.18	24.75 ± 0.81	45.25 ± 2.37	65.95 ± 1.33
Gurzuf	10.24 ± 0.19	35.65 ± 0.91	51.95 ± 1.71	78.15 ± 1.80
Lezginka	8.97 ± 0.23	25.20 ± 0.79	68.45 ± 1.62	85.50 ± 1.44
Love	9.42 ± 0.20	32.85 ± 0.84	53.50 ± 1.38	70.20 ± 1.92
Komsomolsky Ogonyok	7.92 ± 0.14	21.25 ± 0.64	50.17 ± 1.91	73.25 ± 2.68
Rosa Kordesii group				
Dortmund	7.58 ± 0.17	6.75 ± 1.66	90.42 ± 2.94	117.00 ± 1.64
Gutsulochka	6.80 ± 0.17	25.60 ± 0.98	44.20 ± 2.30	80.20 ± 1.80
Letniye Zvyozdy	8.20 ± 0.17	42.15 ± 1.60	41.00 ± 1.95	64.80 ± 1.36

beginning of vegetation with Lezginka and Komsomolsky Ogonyok being the latest-blooming in the Grandiflora group.

Study of the phenological phases demonstrated that Komsomolsky Ogonyok entered the budding phase later than all the other grandiflora flowers, so Komsomolsky Ogonyok, by this indicator, was closer to the Kordesii. Moreover, Komsomolsky Ogonyok's second flowering started later than that of other representatives of the Grandiflora group and was similar to that of the Kordesii roses. Its phenorhythmic¹ turned out to

be closest to that of the Dortmund cultivar from the Kordesii group (see Table 2).

Such morphological characters as flower diameter, number of petals, peduncle length and rose-bush height were investigated. The cultivars were compared separately by garden groups according to the classification of the World Federation of Rose Societies (WFRS). The analysis demonstrated that Komsomolsky Ogonyok also significantly differed from the other cultivars (Table 3): it had statistically significant differences ($p < 0.05$) for all characters with Queen Elizabeth and Lezginka, as well as for three characters (flower diameter, number of petals and bush height) with Korallovy Syurpriz. Comparison against Gurzuf and Love also showed statistically

¹ Phenorhythmic describes the phenological rhythms of growth and development of organisms, adapted to the seasonal rhythm of environmental factors and expressed in a clear alternation of phenological phases. The alternation of phenophases is illustrated by phenospectra (Dedu, 1989).

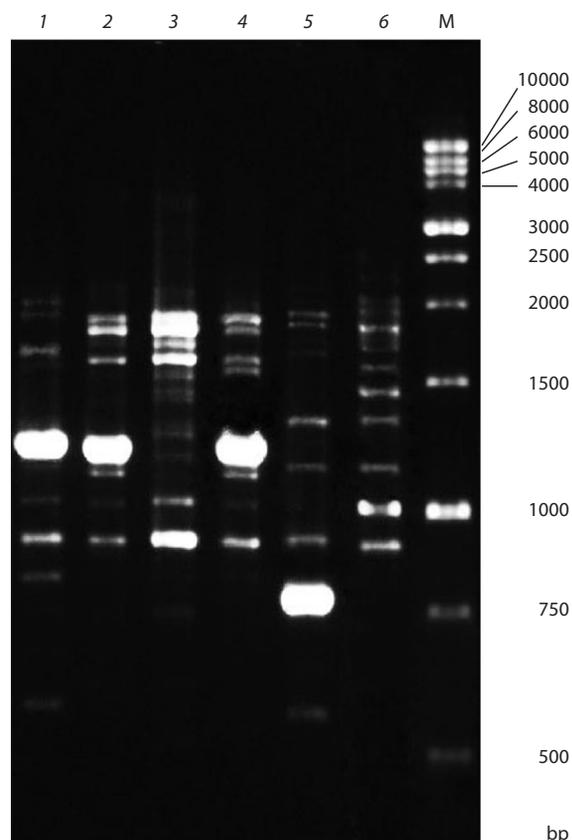


Fig. 1. Electrophoregram of PCR products obtained by DNA amplification with ISSR primer HB12 – (CAC)₃GC.

Tracks with designation of samples: 1 – Komsomolsky Ogonyok, 2 – Queen Elizabeth, 3 – Korrallovy Syurpriz, 4 – Lezginka, 5 – Gurzuf, 6 – Love; track M – DNA marker.

significant differences ($p < 0.05$) for two characters (flower diameter, number of petals). The found pheno- and morphological characters gave us grounds to assume that Komsomolsky Ogonyok should not be referred to the Grandiflora group because its small flower diameter and small number of petals made it closer to the Kordesii roses.

However, comparing Komsomolsky Ogonyok's morphometric characters to those of the Kordesii revealed statistically significant differences as well ($p < 0.05$): it differed from Dortmund and Letniye Zvyozdy by the number of petals, flower stalk and bush heights, and from Gutsulochka by flower diameter, number of petals and bush height.

So, the analyzed phenological phases and morphometric characters showed that Komsomolsky Ogonyok differed from the cultivars of both groups. To assess the degree of kinship, ISSR analysis was employed.

DNA amplification with six ISSR primers identified 122 PCR fragments ranging from 250 to 3,000 bp in length, including 109 polymorphic ones. The number of amplification fragments ranged from 18 (markers HB12 and UBC834) to 23 (17899B) (Fig. 1). The level of polymorphism detected by a single primer ranged from 77.8 % (HB12) to 94.4 % (UBC855) and averaged to 91.42 %.

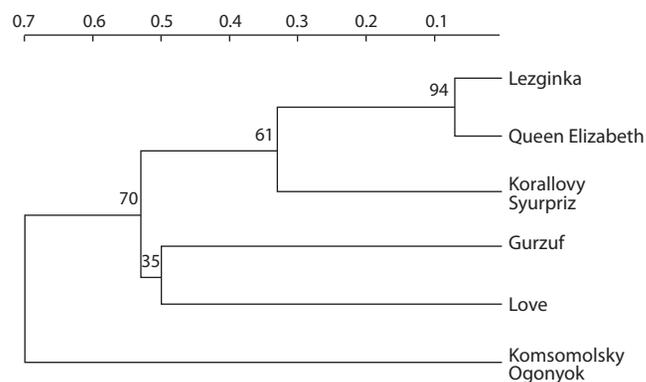


Fig. 2. Dendrogram built with the Neighbor-Joining algorithm based on PCR spectra data of cultivars from the Grandiflora group.

Numbers at nodes show the level of statistical support (100 bootstrap replicates). Numbers at the top show genetic distance.

Based on the obtained results, the samples in the study were divided into three clusters (Fig. 2): Cluster I including Lezginka, Queen Elizabeth and Korrallovy Syurpriz; Cluster II including Gurzuf and Love. Komsomolsky Ogonyok was found to be the most distant from the other cultivars and formed a separate Cluster III.

Queen Elizabeth, bred in the middle of the 20th century, has been widely used for breeding new cultivars, and most likely was a parental one for the Lezginka roses bred in the Nikitsky Botanical Garden in 2005, which was evidenced by the statistically significant² (>90) genetic distance (0.1).

At the next stage of our study, we compared Komsomolsky Ogonyok with the cultivars from the Kordesii group. This comparison revealed 103 amplified fragments ranging from 350 to 2,000 bp in length, including 97 polymorphic ones. The total number of identified fragments ranged from 15 (UBC855) to 19 (HB12) (Fig. 3). The level of polymorphism detected by a single primer ranged from 88.9 % (17899B) to 100 % (M2) and averaged to 94.25 %.

The comparison showed no affinity between the Komsomolsky Ogonyok and Kordesii roses (Fig. 4). The highest kinship score was found for Dortmund and Letniye Zvyozdy. Apparently, the Dortmund cultivar was a parental one in this pair. Gutsulochka was also found to be related to the Dortmund and Letniye Zvyozdy cultivars, but their kinship was less pronounced and its statistical significance was somewhat lower, so these three cultivars of the Kordesii group were brought into Cluster IV.

Thus, the results of molecular analysis as well as investigation of the pheno- and morphometric parameters of the roses in the two groups have shown that Komsomolsky Ogonyok stands out from the members of both groups.

Discussion

The classification system for garden roses has been refined and undergone various, sometimes diametrically opposed, changes over the past 50 years; until the 1970s, the world's

² Confidence measure for a node of interest of >70 (95 % CI) is considered as highly reliable.

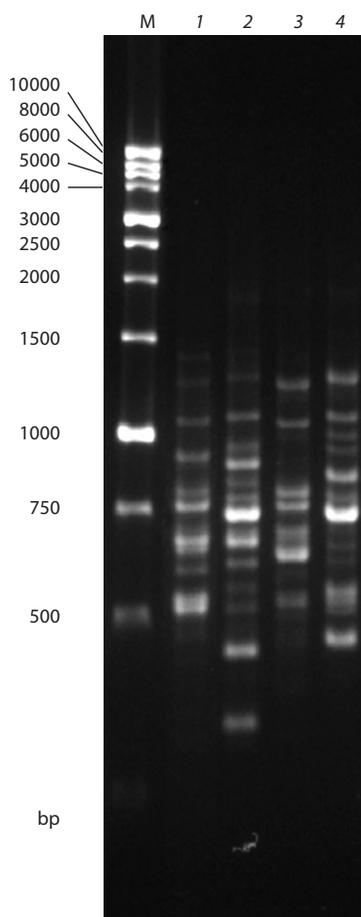


Fig. 3. Electrophoregram of PCR products obtained by DNA amplification with ISSR primer UBC855 – (AC)₈YT.

Tracks with designation of samples: 1 – Dortmund, 2 – Letniye Zvyozdy, 3 – Komsomolsky Ogonyok, 4 – Gutsulochka; track M – DNA marker.

garden roses amounted to approximately 25,000 cultivars subdivided into 30 garden groups (Bylov et al., 1972). In the 1980s (Klimenko, Rubtsova, 1986), foreign specialists in rose breeding and varietal evaluation reduced the number of garden groups to 16. One of the most prominent examples of this merger was the Rambler (climbing roses) group now uniting the Multiflora and Wichuraiana roses. Initially, these groups had clear differences, as they were bred from two different species belonging to the same *Synstylae* section, *Rosa multiflora* Thunb. and *R. wichuraiana*, respectively. But further crosses between these groups resulted in cultivars with the characters common to the original species, and obvious distinction between the groups disappeared. Remarkably, modern molecular genetic studies (Cui et al., 2020) still distinct the original species.

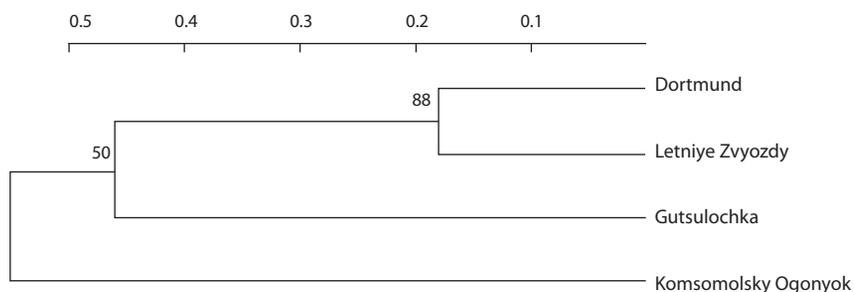


Fig. 4. Dendrogram built based on ISSR-PCR data for cultivars from the Kordes Roses group. Numbers at nodes show the level of statistical support (100 bootstrap replicates). Numbers at the top show genetic distance.

The genetic studies to clarify taxonomic, phylogenetic relationships in roses are intensively developing. They include building genetic, genomic and transcriptomic tools to investigate the molecular mechanisms underlying the creation of some rose species (Bendahmane et al., 2013; Duta-Cornescu et al., 2017; Li et al., 2018). In particular, the genetic kinship of the Taif roses to some rose genotypes (*Rosa* sp.) was assessed based on random amplified polymorphic DNA and simple sequence repeat markers *inter* and *simple* (El-Assal et al., 2014), and Chinese researchers identified the transcripts common to the rose family, which should help clarify the phylogenetic relationships in it (Li et al., 2018).

Rose introduction and breeding has had a long and complex history since the plant has been crossbred in completely different regions of the world, such as Europe, Asia and the Middle East. Domesticating the rose, breeders have concentrated on several characters affecting flower quality such as periodic flowering; terry flowers; petal coloration and fragrance (Bendahmane et al., 2013). So, stimulation of flowering, flower longevity, and creation of novelty in flower structure, color range and fragrances are the main objectives of ornamental plant breeding today. New genome editing techniques offer new opportunities to study the rose's gene function and develop new cultivars for the floriculture industry (Giovannini et al., 2021). Knowing the genetic structure of a species, genus or family allows for more rational use of the available gene pool by optimizing initial breeding forms selection.

Over the past two decades, the molecular basis of floral fragrance as well as its genetic inheritance have been studied in the rose, providing useful information for both researchers and manufacturers (Yan et al., 2014; Shi, Zhang, 2022). A complex research carried out by French, Chinese and German scientists has led to the sequencing of the genome of the tea rose (*Rosa chinensis*), a progenitor species for many modern cultivars, and the genes presumably responsible for its remontancy have been discovered. They have also found that the synthesis of the volatiles giving the flower its fragrance and the pigments responsible for its color are coordinated by the same tandem of a protein and a non-coding microRNA (Raymond et al., 2018).

Conclusion

As has been mentioned above, the world's entire assortment of garden and park roses exceeds 40,000 cultivars subdivided into 36 garden groups. Many of them have no clear confirmation of their origin, e. g., the only indication of the Dortmund cultivar's origin in the catalogs is Seedling × *R. kordesii*. In this situation, creating scientific collections of roses, their gene pools, and passportization of their cultivars becomes of great importance, because intervarietal rose hybridization within the same garden group has largely exhausted itself. It is necessary to search for new cultivars, primarily among the Tea-Hybrid, Floribunda and Grandiflora groups. Apart from high ornamentality, these cultivars are to have distinguishing morphological, rhythmological and resistance characters for use in breeding as paternal and maternal forms.

In the present study, rose cultivars from the bioresource scientific collection of CSBG SB RAS (USU 440534) belonging to the Grandiflora and Kordesii garden groups were investigated. The study has demonstrated that the Komsomolsky Ogonyok cultivar has different molecular genetic, pheno- and morphological characters than those of the Grandiflora group it belongs to. At the same time, it is not closely related to the investigated Kordesii cultivars, despite being close to them in flower size and phenorhythmics.

Such cultivars that are phenotypically different from others and possess a number of valuable characters – primarily winter hardiness, resistance to fungal diseases, and decorative, remontant and prolonged flowering – are valuable breeding material. Our research has shown it is Komsomolsky Ogonyok that can be recommended for breeding rose cultivars for regions with harsh climatic conditions. The polymorphism revealed from ISSR marking data can be used for molecular genetic passportization, which is a necessary step for accounting and conservation of the gene pool of valuable cultivars. The advantage of this approach is that the ISSR technique is polylocus, has a large number of PCR amplification products and does not require sequencing in polyacrylamide gels.

References

- Adritskaya N.A., Kapelyan A.I. Comparative assessment of winter hardiness and decorative properties of various varieties of roses in the rose garden of the Botanical Garden of Peter the Great. *Izvestiya Sankt-Peterburgskogo Gosudarstvennogo Agrarnogo Universiteta = Proceedings of the Saint Petersburg State Agrarian University*. 2022;2(67):48-58. DOI 10.24412/2078-1318-2022-2-48-58 (in Russian)
- Amom T., Nongdam P. The use of molecular marker methods in plants: a review. *Int. J. Cur. Res. Rev.* 2017;9(17):1-7. DOI 10.7324/IJCRR.2017.9171
- Annotated Catalog of Ornamental Plants from the Collection of the Nikita Botanical Garden. Vol. I. Collections of garden roses, clematis, and lilacs. Simferopol: ARIAL Publ., 2018 (in Russian)
- Bardakova S.A. Influence of adverse climatic conditions on the growth and development of garden roses in the Stavropol Botanical Garden. *Vestnik APK Stavropolya = Agricultural Bulletin of the Stavropol Region*. 2017;1(25):120-122 (in Russian)
- Beideman I.N. Methodology for Studying the Phenology of Plants and Plant Communities. Novosibirsk: Nauka Publ., 1974 (in Russian)
- Bendahmane M., Dubois A., Raymond O., Bris M.L. Genetics and genomics of flower initiation and development in roses. *J. Exp. Bot.* 2013;64(4):847-857. DOI 10.1093/jxb/ers387
- Bylov V.N., Shtanko I.I., Yudinseva E.V., Mikhailov N.L. Roses. Summary of introduction. Moscow: Nauka Publ., 1972 (in Russian)
- Cairns T. (Ed.) Modern Roses XI: The world encyclopedia of roses. New York: Acad. Press, 2007
- Cui W.H., Zhong M.C., Du X.Y., Qu X.J., Jiang X.D., Sun Y.B., Wang D., Chen S.Y., Hu J.Y. The complete chloroplast genome sequence of a rambler rose, *Rosa wichuraiana* (Rosaceae). *Mitochondrial DNA Part B*. 2020;5(1):252-253. DOI 10.1080/23802359.2019.1700198
- Dedu I.I. Ecological Encyclopedic Dictionary. Kishinev, 1989 (in Russian)
- Dorogina O.V., Zhmud E.V. Molecular-genetic methods in plant ecology. *Contemporary Problems of Ecology*. 2020;13(4):333-345. DOI 10.1134/S1995425520040058
- Doyle J.J., Doyle J.L. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bul.* 1987;19(1):11-15
- Duta-Cornescu G., Pavlusenco C.-E., Pojoga D.M., Negulici M.E., Constantin N., Simon-Gruitu A. Genetic analysis of some roses cultivars appropriate for S-E Romania climate using PCR-ISSR technology. *AgroLife Sci. J.* 2017;6(1):69-74
- El-Assal S.E.D., El-Awady M.A., El-Tarras A., Shehab G. Assessing the genetic relationship of Taif rose with some rose genotypes (*Rosa* sp.) based on random amplified polymorphic DNA, inter simple sequence repeat and simple sequence repeat markers. *Am. J. Biochem. Biotechnol.* 2014;10(1):88-98. DOI 10.3844/ajbbsp.2014.88.98
- Fomina T.I. Biological Characteristics of Ornamental Plants of Natural Flora in Western Siberia. Novosibirsk: Acad. Publ. House "Geo", 2012 (in Russian)
- Giovannini A., Laura M., Nesi B., Savona M., Cardi T. Genes and genome editing tools for breeding desirable phenotypes in ornamentals. *Plant Cell Rep.* 2021;40(3):461-478. DOI 10.1007/s00299-020-02632-x
- Gorodnyaya E.V. A promising assortment of roses for use in landscaping and breeding in the foothill zone of Crimea. *Uchenye Zapiski Tavricheskogo Natsionalnogo Universiteta imeni Vernadskogo. Seriya Biologiya, Khimiya = Proceedings of the Vernadsky Taurida National University. Series: Biology, Chemistry*. 2014;27(5):29-37 (in Russian)
- Haynes W. Student's t-test. In: Dubitzky W., Wolkenhauer O., Cho K.H., Yokota H. (Eds.) Encyclopedia of Systems Biology. New York: Springer, 2013;2023-2025. DOI 10.1007/978-1-4419-9863-7_1184
- Kapelyan A.I. Grafted and own-rooted roses in the Saint Petersburg Botanical Garden. *Collection of Scientific Works of the Nikita State Botanical Garden*. 2017;145:271-274 (in Russian)
- Klimenko Z.K. The results of long-term work (1824–2010) on the breeding of garden roses in the Nikita Botanical Garden. *Byulleten Gosudarstvennogo Nikitskogo Botanicheskogo Sada = Bulletin of the Nikita State Botanical Garden*. 2010;100:49-55 (in Russian)
- Klimenko Z.K., Rubtsova E.L. Roses. Kyiv: Naukova dumka Publ., 1986 (in Russian)
- Klimenko Z.K., Vasilyeva O.Yu., Zorina E.V., Dzyuba O.V. Ecological and geographical testing of garden roses in three climatic zones. *Samarskiy Nauchnyi Vestnik = Samara Journal of Science*. 2019;8(1):36-42. DOI 10.24411/2309-4370-2019-11105 (in Russian)
- Li S., Zhong M., Dong X., Jiang X., Xu Y., Sun Y., Cheng F., Li D.Z., Tang K., Wang S., Dai S., Hu J.Y. Comparative transcriptomics identifies patterns of selection in roses. *Plant Biol.* 2018;18(1):371. DOI 10.1186/s12870-018-1585-x
- Methodology of State Variety Testing of Agricultural Crops. Iss. 6. Ornamental crops. Moscow: Kolos Publ., 1968 (in Russian)
- Nei M., Li W.-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA*. 1979;76(10):5269-5273. DOI 10.1073/pnas.76.10.5269
- Pashina M.V. Rhythms of growth and development of garden roses in the Irtysh forest-steppe. *Vestnik IrGSHA*. 2011;6(44):110-116 (in Russian)
- Plugatar Yu.V., Klimenko Z.K., Plugatar S.A., Zykova V.K., Kravchenko I.N. Rambling roses in the landscape of the southern coast of the Crimea: historical traditions of their use. *Acta Hort.* 2018;1201:49-55. DOI 10.17660/ActaHortic.2018.1201.88
- Plugatar Yu.V., Klimenko Z.K., Zykova V.K., Plugatar S.A. Methods and results of roses breeding from different garden groups in the south of Russia. *Acta Hort.* 2019;1255:31-34. DOI 10.17660/ActaHortic.2019.1255.6
- Raymond O., Gouzy J., Just J., Badouin H., Verdenaud M., Lemainque A., Vergne P., Moja S., Choisine N., Pont C., ... Liu C., Le Bris M., Salse J., Baudino S., Benhamed M., Wincker P., Bendahmane M. The *Rosa* genome provides new insights into the domestication of modern roses. *Nat. Genet.* 2018;50(6):772-777. DOI 10.1038/s41588-018-0110-3

- Shi S., Zhang Z. Genetic and biochemical aspects of floral scents in roses. *Int. J. Mol. Sci.* 2022;23(14):8014. DOI 10.3390/ijms23148014
- Suprun N.A. Comprehensive assessment of the collection of rose varieties of the VSSPU Botanical Garden. In: Proceedings on the Introduction and Acclimatization of Plants. Izhevsk, 2021;525-529 (in Russian)
- Tyshchenko E.L. Promising garden groups of roses for use in landscaping in southern Russia. In: Scientific Works of the North Caucasian Zonal Research Institute for Horticulture and Viticulture. 2015;7: 67-72 (in Russian)
- Van de Peer Y., Wachter R.D. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput. Appl. Biosci.* 1994; 10(5):569-570
- Vasilyeva O.Yu. Introduction of Roses in Western Siberia. Novosibirsk: Nauka Publ., 1999 (in Russian)
- Vasilyeva O.Yu., Dorogina O.V., Yudanova S.S., Plugatar S.A., Klimenko Z.K. Identifying the rose varieties and natural forms using ISSR-markers. *BIO Web Conf.* 2020;24:00091. DOI 10.1051/bioconf/20202400091
- Yan H., Zhang H., Chen M., Jian H., Baudino S., Caissard J.C., Bendahmane M., Li S., Zhang T., Zhou N., Qiu X., Wang Q., Tang K. Transcriptome and gene expression analysis during flower blooming in *Rosa chinensis* 'Pallida'. *Gene.* 2014;540(1):96-103. DOI 10.1016/j.gene.2014.02.008

ORCID

S.S. Yudanova orcid.org/0000-0001-7547-0099
O.V. Dorogina orcid.org/0000-0001-5729-3594
O.Yu. Vasilyeva orcid.org/0000-0003-0730-3365

Acknowledgements. The study was carried out as part of Complex Program for Fundamental Scientific Research of the Siberian Branch of the Russian Academy of Sciences (Project No. AAAA-A21-121011290025-2 "Analysis of Biodiversity, Conservation and Restoration of Rare and Resource Plant Species Using Experimental Methods"), using the materials provided by CSBG SB RAS's Collections of Living Plants in Open and Protected Grounds, USU 440534.

Conflict of interest. The authors declare no conflict of interest.

Received June 28, 2023. Revised December 11, 2023. Accepted December 11, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Manifestation of agronomically valuable traits in the progeny of a sorghum mutant carrying the genetic construct for RNA silencing of the γ -kafirin gene

L.A. Elkonin¹✉, N.V. Borisenko¹, T.E. Pylaev^{2,3}, O.A. Kenzhegulov¹, S.Kh. Sarsenova¹, N.Yu. Selivanov², V.M. Panin¹

¹ Federal Center of Agriculture Research of the South-East Region, Saratov, Russia

² Institute of Biochemistry and Physiology of Plants and Microorganisms, Saratov Federal Scientific Center of the Russian Academy of Sciences, Saratov, Russia

³ Saratov State Medical University named after V.I. Razumovsky, Saratov, Russia

✉ elkonin@gmail.com

Abstract. Improving the nutritional value of grain sorghum, a drought- and heat-tolerant grain crop, is an important task in the context of global warming. One of the reasons for the low nutritional value of sorghum grain is the resistance of its storage proteins (kafirins) to proteolytic digestion, which is due, among other things, to the structural organization of protein bodies, in which γ -kafirin, the most resistant to proteases, is located on the periphery, encapsulating more easily digested α -kafirins. The introduction of genetic constructs capable of inducing RNA silencing of the γ -kafirin (*gKAF1*) gene opens up prospects for solving this problem. Using *Agrobacterium*-mediated genetic transformation of immature embryos of the grain sorghum cv. Avans we have obtained a mutant with improved digestibility of endosperm proteins (up to 92 %) carrying a genetic construct for RNA silencing of the *gKAF1* gene. The goal of this work was to study the stability of inheritance of the introduced genetic construct in T₂–T₄ generations, to identify the number of its copies, as well as to trace the manifestation of agronomically valuable traits in the offspring of the mutant. The mutant lines were grown in experimental plots in three randomized blocks. The studied lines were characterized by improved digestibility of kafirins, a modified type of endosperm, completely or partially devoid of the vitreous layer, an increased percentage of lysine (by 75 %), reduced plant height, peduncle length, 1000-grains weight, and grain yield from the panicle. In T₂, a line with monogenic control of GA resistance was selected. qPCR analysis showed that in different T₃ and T₄ plants, the genetic construct was present in 2–4 copies. In T₃, a line with a high digestibility of endosperm proteins (81 %) and a minimal decrease in agronomically valuable traits (by 5–7 %) was selected.

Key words: *Sorghum bicolor*; transgenic plants; RNA silencing; kafirins; qPCR; grain quality; endosperm texture.

For citation: Elkonin L.A., Borisenko N.V., Pylaev T.E., Kenzhegulov O.A., Sarsenova S.Kh., Selivanov N.Yu., Panin V.M. Manifestation of agronomically valuable traits in the progeny of a sorghum mutant carrying the genetic construct for RNA silencing of the γ -kafirin gene. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(1):63-73. DOI 10.18699/vjgb-24-08

Проявление селекционно-ценных признаков в потомстве мутанта сорго, несущего генетическую конструкцию для РНК-сайленсинга гена γ -кафирина

Л.А. Эльконин¹✉, Н.В. Борисенко¹, Т.Е. Пылаев^{2,3}, О.А. Кенжегулов¹, С.Х. Сарсенова¹, Н.Ю. Селиванов², В.М. Панин¹

¹ Федеральный аграрный научный центр Юго-Востока, Саратов, Россия

² Институт биохимии и физиологии растений и микроорганизмов, Федеральный исследовательский центр «Саратовский научный центр Российской академии наук», Саратов, Россия

³ Саратовский государственный медицинский университет им. В.И. Разумовского Министерства здравоохранения Российской Федерации, Саратов, Россия

✉ elkonin@gmail.com

Аннотация. Улучшение питательной ценности зернового сорго – засухоустойчивой и жаростойкой зерновой культуры, служащей источником кормового и продовольственного зерна во многих засушливых регионах мира, – важная задача селекции в условиях глобального потепления климата. Одной из причин низкой питательной ценности зерна сорго является устойчивость его запасных белков (кафиринов) к протеолитическому расщеплению, обусловленная, в том числе структурной организацией белковых телец, в которых наиболее устойчивый к действию протеаз γ -кафирин располагается по периферии, инкапсулируя более легко перевариваемые α -кафирины. Введение генетических конструкций, способных к индукции РНК-сайленсинга гена γ -кафирина (*gKAF1*), открывает перспективы для решения этой проблемы. Нами у сорта зернового сорго Аванс

посредством агробактериальной трансформации с использованием штамма, несущего в составе T-ДНК генетическую конструкцию для РНК-сайленсинга гена *gKAF1*, получен мутант с улучшенной перевариваемостью белков эндосперма (до 92 %). Цель настоящей работы – изучение стабильности наследования введенной генетической конструкции в поколениях T₂–T₄, установление числа ее копий, а также проявления важнейших селекционно-ценных признаков у потомства полученного мутанта. Мутантные линии выращивали на экспериментальном участке Федерального аграрного научного центра Юго-Востока в трех рендомизированных блоках. Исследуемые линии характеризовались улучшенной перевариваемостью кафиринов, модифицированным типом эндосперма, полностью или частично лишенным стекловидного слоя, повышенным процентным содержанием лизина (на 75 %), сниженной высотой растений, длиной подметельчатого междоузлия, массой 1000 зерен и урожаем зерна с метелки. В поколении T₂ отобрана линия с моногенным контролем устойчивости к глюфосинату аммония. Анализ qPCR показал, что у разных растений из поколений T₃ и T₄ генетическая конструкция присутствует в двух-четыре копии. В поколении T₃ отобрана линия с высокой перевариваемостью белков эндосперма (81 %) и минимальным снижением селекционно-ценных признаков (на 5–7 %).

Ключевые слова: *Sorghum bicolor*; трансгенные растения; РНК-сайленсинг; кафирины; qPCR; качество зерна; текстура эндосперма.

Introduction

RNA interference (RNAi) is one of the most important natural mechanisms for regulating gene expression and antiviral cell defense. As is known, the mechanism of RNAi is based on the degradation of single-stranded mRNA in the presence of complementary short RNA, which leads to disruption of protein synthesis and silencing of the expression of the corresponding gene. These short interfering RNAs (siRNAs), 20–25 nucleotides long, are transcribed from natural DNA sequences present in the genome or artificially created genetic constructs encoding hairpin RNAs (hpRNAs) (Guo et al., 2016; Muhammad et al., 2019). This design consists of sense and antisense sequences of the target gene mRNA in the form of inverted repeats, which are separated by a spacer sequence. A splicable intron is often used as a spacer in such genetic designs because it improves the efficiency of RNA silencing in plants (Smith et al., 2000).

The sense and antisense sequences in the transcribed RNA are complementary to each other and form hpRNA, which is processed by Dicer-like proteins (DCLs). DCL proteins generate siRNAs from a precursor, hpRNA. One strand of the siRNA duplex is incorporated into the Argonaute protein (AGO), forming the RNA-induced silencing complex (RISC). The siRNA molecule directs RISC to the complementary region of the single-stranded RNA, after which AGO cleaves the target mRNA (Zhuravlyov et al., 2022; Bharathi et al., 2023).

RNAi technology is frequently used in functional genomics and genetic engineering of plants, since it makes it possible to create mutants that are resistant to biotic and abiotic stress factors, as well as to regulate the expression of genes involved in the synthesis and catabolism of important metabolites and reserve nutrients, including proteins, carbohydrates and lipids (Bharathi et al., 2023). This approach has been used to obtain mutants with an altered spectrum of seed storage proteins in cultivated cereal species, including maize, wheat, rice, and sorghum (Elkonin et al., 2016a).

The use of RNAi technology for regulating accumulation of seed storage proteins is of particular importance for sorghum, since it is believed that one of the reasons for the relatively low nutritional value of sorghum grain is the resistance of its storage proteins (kafirins) to proteolytic digestion, due, among other things, to the structural organization of protein bodies, in which the one most stable to the action of proteases, γ -kafirin,

is located at the periphery, encapsulating the more easily digestible α -kafirins (Oria et al., 2000; de Mesa-Stonestreet et al., 2010; Duressa et al., 2018). It has been shown that suppression of the synthesis of different classes of kafirins leads to modification of the structure of endosperm protein bodies, a decrease in the resistance of kafirins to protease digestion, and compensatory synthesis of other proteins with a higher lysine content (da Silva et al., 2011a, b; Kumar et al., 2012; Grootboom et al., 2014; Elkonin et al., 2016b).

Previously, through *Agrobacterium*-mediated genetic transformation, a genetic construct capable of inducing RNA silencing of the γ -kafirin gene (*gKAF1*), which is a part of the T-DNA vector pNRKAFSIL (Elkonin et al., 2016b), was introduced into the genome of the commercial grain sorghum cultivar Avans. In one of the T₀ plants, all kernels had a floury endosperm type and were characterized by a high level of kafirins digestibility (up to 93 %), which was significantly higher than the level of digestibility in the cv. Avans (57 %) (Elkonin et al., 2021). T₁ plants inherited these traits, as well as resistance to ammonium glufosinate caused by the *bar* gene, which was also present in the T-DNA of the pNRKAFSIL vector (Borisenko et al., 2022).

The purpose of this work was to study the stability of inheritance of the introduced genetic construct in T₂–T₄ generations, its copy number, as well as the manifestation of the most important agronomically valuable traits in the offspring of the mutant, including the amino acid composition of flour and the digestibility of its proteins in an *in vitro* system.

Materials and methods

The original mutant with high digestibility of grain proteins (Avans-1/18) was obtained in an experiment on *Agrobacterium*-mediated genetic transformation of the new commercial cultivar Avans of grain sorghum (*Sorghum bicolor* (L.) Moench) using *Agrobacterium tumefaciens* strain GV3101, carrying the binary vector pNRKAFSIL (Elkonin et al., 2021). This vector contains a genetic construct consisting of a fragment (309 bp) of the γ -kafirin gene (GenBank acc. No. M73688) in direct and inverted orientation, which are separated by the maize *ubi1*-intron sequence (Fig. 1) (Elkonin et al., 2016b). This construct should suppress *gKAF1* expression via RNA interference. The T-DNA of this vector also includes the *bar* gene under the control of the *nos*-promoter, which determines resistance to ammonium glufosinate (GA).

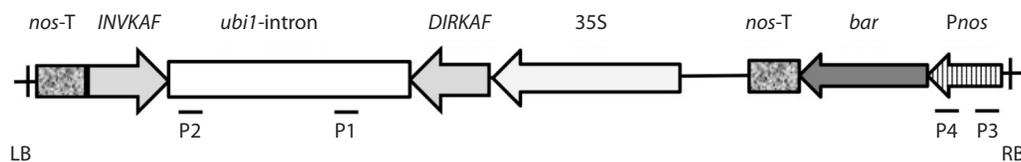


Fig. 1. T-DNA of the pNRKAFSIL vector (Elkonin et al., 2016b).

DIRKAF, *INVKAF* – fragments of the sorghum γ -kafirin gene (GenBank Acc. No. M73688) in direct and inverted orientation; maize *ubi1*-intron; 35S – 35S-promoter of cauliflower mosaic virus; *Pnos* – *nos*-promoter; *nos-T* – *nos*-terminator; P1–P4 – positions of primers used in PCR analyses.

The progeny of the RNAi mutant Avans-1/18 was grown in a growth chamber (photoperiod: 16 hours day/8 hours night; t° : 28/22 $^\circ\text{C}$), or in a greenhouse, and in experimental plots in the field of the Federal Center of Agriculture Research of the South-East Region (Saratov, Russia).

To study inheritance of the genetic construct for silencing, panicles of transgenic plants were carefully isolated with parchment bags before flowering. In addition, they were crossed as paternal parents with lines with cytoplasmic male sterility (CMS) A2 KVV-181, A2 KVV-114 and A2 O-1237.

Assessment of resistance to ammonium glufosinate.

To study the resistance of the progeny of transgenic plants to GA, the kernels were sterilized with Domestos, then with HgCl_2 (0.1 %), after which they were soaked in a GA solution (2.5 mg/l) for 20 hours. After that, the embryos were removed from the soaked kernels and placed on a hormone-free MS nutrient medium containing filter-sterilized GA (2.5 mg/l). The embryos were cultivated in a growth chamber (photoperiod: 16 hours day/8 hours night; t° 26–28 $^\circ\text{C}$) for three weeks. Under this regime, seedlings carrying the *bar* gene developed normally and reached 10 cm in three weeks, while the development of seedlings not carrying the *bar* gene was inhibited at the coleoptile growth stage.

Endosperm texture. The texture of the endosperm was determined on transverse sections of mature kernels. The following endosperm types were distinguished: modified, which included the floury type, floury with interspersed vitreous endosperm and floury with a thin, often “blurred” rim of the vitreous layer, and regular endosperm with a thick vitreous layer.

PCR analysis. Genomic DNA was isolated from leaves using a modified CTAB method. The presence of a genetic construct for RNA silencing was checked using PCR analysis with primers that amplified a 588-bp fragment of the maize *ubi1*-intron (P1 (5'→3') tgtcttggttgatgatgtggtc; P2 (5'→3'): gcgatgaaggcagggctaaa), which is an important component of RNAi genetic construct, ensuring its stability and functional efficiency (Smith et al., 2000), and the fragment of the *nos*-promoter that controls the *bar* marker gene, 201 bp long (P3 (5'→3'): tgagacttaattggataccgagg; P4 (5'→3'): ttggaac tgacagaaccgcaac) (primer positions are indicated in Fig. 1). PCR was carried out using DNA amplifiers MasterCycler (Eppendorf, Germany) and T100 (Bio-Rad, USA). PCR conditions were as follows: for the *nos*-promoter: 95 $^\circ\text{C}$ (2 min); 40 cycles [95 $^\circ\text{C}$ (30 s), 64 $^\circ\text{C}$ (30 s); 72 $^\circ\text{C}$ (1 min 10 s)]; 72 $^\circ\text{C}$ (7 min). For *ubi1*-intron: 95 $^\circ\text{C}$ (2 min); 40 cycles [95 $^\circ\text{C}$ (1 min), 56 $^\circ\text{C}$ (1 min), 72 $^\circ\text{C}$ (1 min 30 s)]; 72 $^\circ\text{C}$ (10 min). The amplified fragments were visualized by electrophoresis

in a 2.0 % agarose gel. PCR analyses of each sample were performed in two replications.

qPCR analysis. The copy number of the genetic construct for RNAi was determined using quantitative PCR by amplifying a 119 bp fragment of sorghum anthranilate phosphoribosyl transferase gene (*APRT*, Sobic.002G303300), selected as a reference gene (Casu et al., 2012; Wang et al., 2021), and a 201 bp fragment of the *nos*-promoter. The reaction mixture contained 2 μl of genomic DNA (10 ng/ μl), 10 μl of a ready-made amplification reaction mixture containing the intercalating dye SYBR Green (2X HS-qSYBR-blue, Biolabmix, Russia) and 0.4 pmol of each primer; the total volume of the reaction mixture was 20 μl ; the number of replications was three. PCR mode: 1 cycle of 95 $^\circ\text{C}$ (2 min), then 40 cycles of [95 $^\circ\text{C}$ (15 s), 60 $^\circ\text{C}$ (20 s) + detection on the FAM channel, 72 $^\circ\text{C}$ (20 s)]. qPCR was performed using a LightCycler 96 real-time PCR instrument (Roche, Switzerland). The primers for the *APRT* gene were as follows (F (5'→3'): tgacacattccc aacctcaa and R (5'→3'): atctctccctgagtggca) (Wang et al., 2021); primers for the *nos*-promoter are described above. The concentration of genomic DNA and primers was determined using a Nanodrop One C UV spectrophotometer (Thermo Fisher Scientific, USA) by measuring absorbance at wavelengths of 230/260/280 nm. Data analysis and determination of PCR efficiency were performed using the certified LightCycler® 96 Software, version 1.1.0.1320.

Analysis of total protein content in grain. The total grain protein content of plants from T₁ and T₂ was analyzed using the Kjeldahl method.

Digestibility of grain proteins. To study the digestibility of grain proteins, pepsin treatment of whole-grain flour was used (Aboubacar et al., 2001; Nunes et al., 2004; Wong et al., 2009). A sample of flour (20 mg) was incubated in 5 ml of a 0.15 % pepsin solution (Carl Roth, Germany) (Art.-Nr. KK38.1, 2000 FIP-U/g) in 0.1 M potassium phosphate buffer (pH 2.0) for 120 min at 37 $^\circ\text{C}$ with occasional shaking.

To quantify digestibility, a method based on scanning the electrophoretic spectra of proteins obtained in SDS-PAGE was used (Aboubacar et al., 2001; Wong et al., 2009; Elkonin et al., 2013). Proteins were isolated from digested and control samples using extraction buffer (0.0625 M TRIS·HCl, 2 % SDS, 5 % 2-mercaptoethanol, pH 6.8). Samples were subjected to SDS-PAGE in 12.5 % (w/v) PAGE; 10 μl of extract was applied to each lane. Protein molecular weight markers (14–116 kDa) (Thermo Scientific) were used as standards. Gels were stained with Coomassie Brilliant Blue R-250. After electrophoresis, the gels were scanned using ChemiDoc™ (Bio-Rad); protein amounts were determined using Image

Lab 6.1 (Bio-Rad). The digestibility value was calculated as the percentage of the difference between the volume of protein in the control sample and in the sample treated with pepsin to the volume of protein in the control sample. The experiments were performed in two replications.

HPLC analysis of the amino acid composition of flour proteins. To analyze the total content of amino acids, 10 kernels were examined from transgenic plants T₁ 190-1, 190-3, 190-4 and the original cv. Avans. Ground meal samples (15 mg) were hydrolyzed with 1.5 ml of 6N HCl (at t° of 106 °C for 24 h) under nitrogen. Identification of amino acids was carried out using a pre-column modification with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate according to the Waters AccQ-Tag method using a WAT 052880 reagent kit. The analysis was carried out by HPLC on a Knauer Smartline chromatograph using reverse phase chromatography on a Kromasil – 110 C18/2.5 μ km (2 mm \times 150 mm). Detection was performed at λ 248 nm. The injection volume was 20 μ l. Samples were analyzed in triplicate. Quantitative calculation of the amino acid content in sample hydrolysates was carried out using the external standard method – 250 pM of analytical amino acid standard (AAS18 Fluka).

Evaluation of agronomically valuable traits. To study the influence of the RNAi genetic construct for silencing the *gKAF1* gene on the manifestation of agronomically valuable traits, the T₃ and T₄ progeny of four T₁ transgenic plants, as well as the original Avans variety, were grown in 4 meter rows in three replications in the experimental field of the Federal Centre of Agriculture Research of the South-East Region. All panicles of plants were carefully bagged with parchment bags before flowering. The following traits were analyzed: plant height, peduncle length, weight of 1000 grains, and grain yield per panicle. In each replication, the average value of the trait was determined in 10–15 plants.

Methods of biological statistics. To assess differences in the *in vitro* digestibility of proteins of the studied samples, dispersion analysis and Duncan's Multiple Range Test were performed using the AGROS software package, version 2.09 (S.P. Martynov, Institute of General Genetics of RAS). Data on the manifestation of agronomically valuable traits were processed by one-way dispersion analysis using the AGROS software package.

Results and discussion

PCR analysis of plants from T₂–T₄ generations with primers to the *ubi1*-intron separating inverted fragments of the *gKAF1* gene in the pNRKAFSIL vector, and to the *nos*-promoter that controls the expression of the *bar* gene, showed the stability of inheritance of the genetic construct during self-pollination. In the T₂ generation, the progeny of four T₁ plants was analyzed, each of which was PCR-positive for both loci studied (Nos. 1-3, 190-1, 190-2, 190-4), as well as the progeny of plant No. 190-3, which was PCR negative for the *ubi1*-intron, but PCR positive for the *nos*-promoter (Elkonin et al., 2021).

It was found that all 12 T₂ plants from the T₁ family No. 1-3 and all 8 T₂ plants from the T₁ family No. 190-4 were PCR positive for both loci tested (Table 1). At the same time, in the T₂ families – the progeny of plants 190-1 and 190-2 – segregation was observed, and PCR-negative plants for both loci were present. No PCR-negative plants were found in the T₃ and T₄

generations. This fact indicates the stability of inheritance of the introduced genetic construct.

PCR analysis of F₁ hybrids between T₁ plants and different CMS lines showed the possibility of transmitting the introduced genetic construct through pollen. It is noteworthy that among the F₁ hybrids with the CMS line A2 KVV-114, one plant that was PCR positive for the *nos*-promoter but PCR negative for the *ubi1*-intron was found. We previously found a similar case among T₁ plants (Elkonin et al., 2021). These facts indicate that in some cases there may be a disturbance of the integrity of the integrated T-DNA and the elimination of the genetic construct for silencing. A significant predominance of PCR-positive plants in T₂, as well as among F₁ hybrids, indicated the presence of several unlinked copies of the genetic construct in T₁ plants.

Resistance to ammonium glufosinate. To further study the inheritance of the genetic construct for RNA silencing, the progeny of a number of plants from T₁–T₃ generations was grown on a nutrient medium containing ammonium glufosinate (2.5 mg/l), resistance to which is determined by the presence of a *bar* gene. In different plants, different segregation patterns were observed: 15:1 (plant No. 182-3), indicating the presence of two copies of the genetic construct; monogenic segregation 3:1 (No. 124-3; 124-3-9; 124-3-10), indicating the presence of one copy of the construct; in the progeny of plants No. 150-15, 124-3-3, 190-4, no segregation was observed, which suggested homozygous nature of these plants (Table 2). Considering that the plant 124-3-3 was obtained from the progeny of plant 124-3, which segregated as a mono-heterozygote (3:1), it is quite obvious that in the plant 124-3-3, the genetic construct for the silencing of the *gKAF1* gene is present in one copy in a homozygous state. No resistant plants were observed in the progeny of plant 190-3, as well as in the original cv. Avans.

Endosperm texture. Analysis of the endosperm texture in kernels of plants from different generations of the RNAi mutant Avans-1/18 showed the inheritance of a modified endosperm type (floury or floury with vitreous inclusions) (Fig. 2), characteristic of the mutant kernels, up to the T₄ generation. In the panicles of T₁ and T₂ plants, due to their heterozygosity, a 3:1 or 15:1 segregation (modified endosperm type: normal vitreous endosperm) was observed, which indicated the presence of one or two copies of the genetic construct (Table 3). In the T₃ and T₄ generations, segregation of the kernels with vitreous endosperm was absent, as a rule, indicating the homozygous state of the introduced genetic construct.

It is noteworthy that in many kernels a combined type of endosperm was observed, in which the floury endosperm was interspersed with sectors of vitreous endosperm, or the vitreous endosperm was present in the form of a thin, often “blurred” peripheral layer (see Fig. 2). The formation of these types of endosperms did not depend on the number of copies of the construct (see below) and, possibly, reflected the peculiar properties of its expression under environmental conditions that change during kernel maturation.

Similar “combined” endosperm types were described previously in other works on the introduction of genetic constructs for silencing kafirin genes (da Silva, 2012; Elkonin et al., 2016b), as well as in recombinant lines obtained by hybridization of the P721Q mutant, characterized by high digestibility

Table 1. Inheritance of the genetic construct for RNA silencing of the *gKAF1* gene in the progeny of the RNAi mutant of grain sorghum Avans-1/18 during self-pollination and in crosses with CMS lines

T ₁ plant	Progeny	<i>ubi1</i> -intron			<i>nos</i> -promoter		
		Total	PCR "+"	PCR "--"	Total	PCR "+"	PCR "--"
Self-pollination							
1-3 (<i>ubi1+</i> ; <i>nos+</i>)	1-3/20 (T ₂)	12	12	–	12	12	–
	197/21 (T ₃)	6	6	–	–	–	–
	153/22 (T ₄)	7	7	–	–	–	–
190-1 (<i>ubi1+</i> ; <i>nos+</i>)	181/21 (T ₂)	19	18	1	19	18	1
190-2 (<i>ubi1+</i> ; <i>nos+</i>)	182/21 (T ₂)	15	12	3	15	12	3
	150/22 (T ₃)	14	14	–	–	–	–
190-3 (<i>ubi1–</i> ; <i>nos+</i>)	183/21 (T ₂)	–	–	–	3	–	3
190-4 (<i>ubi1+</i> ; <i>nos+</i>)	184/21 (T ₂)	8	8	–	8	8	–
	152/22 (T ₃)	11	11	–	–	–	–
F ₁ hybrids with CMS-lines							
A2 KVV-181/T ₁ 190-2		17	13	4	17	13	4
A2 O-1237/T ₁ 190-1		15	12	3	15	12	3
A2 KVV-114/T ₁ 190-1		18	16	2	18	17	1

Table 2. Segregation for resistance to ammonium glufosinate in the progeny of some plants of the Avans-1/18 mutant carrying genetic construct for RNA silencing of the *gKAF1* gene

Plant No.	Number of plants			Ratio	χ^2	<i>p</i>
	total	resistant	sensitive			
T ₁ No. 190-2 (<i>nos+</i>) progeny						
124-3/21 (T ₂)	106	81	25	3:1	0.113	0.75 < <i>p</i> < 0.50
182-3/21 (T ₂)	110	101	9	15:1	0.701	0.50 < <i>p</i> < 0.25
150-15/22 (T ₃)	100	100	–	–	–	–
T ₂ No. 124-3 (<i>nos+</i>) progeny						
124-3-3 (T ₃)	35	35	–	–	–	–
124-3-9 (T ₃)	70	49	21	3:1	0.933	0.50 < <i>p</i> < 0.25
124-3-10 (T ₃)	29	21	6	3:1	0.111	0.75 < <i>p</i> < 0.50
T ₁ No. 190-4 (<i>nos+</i>) progeny						
190-4/21 (T ₁)	30	30	–	–	–	–
T ₁ No. 190-3 (<i>nos–</i>) progeny						
151-8/22 (T ₃)	30	–	30	–	–	–

of kafirins and flouy endosperm, with common sorghum lines with low kafirin digestibility and regular vitreous endosperm (Tesso et al., 2006). Considering that the flouy endosperm type is accompanied by a number of negative agronomic traits (Duressa et al., 2018), lines with such combined endosperm types and improved protein digestibility should have higher breeding value.

Analysis of the number of copies of RNAi genetic construct. To clarify the copy number of the genetic construct for silencing of *gKAF1* in T₃ and T₄ plants, quantitative PCR analysis (qPCR) was carried out with primers to the *nos*-promoter, while the sorghum *APRT* gene, which was previously used in experiments to identify the number of copies of transgenes in different plant species, including sorghum (Casu et al., 2012;



Fig. 2. Cross sections of kernels from the progeny of the RNAi grain sorghum mutant Avans-1/18 (a, b) and the original non-transgenic cv. Avans (c). A thick layer of vitreous endosperm is noticeable in the kernels of the original cultivar (c) and there is a complete absence of vitreous endosperm (a) or a blurred layer of vitreous endosperm in the kernels of transgenic T_1 plants (b).

Table 3. Segregation by the endosperm type in the kernels of some plants from different generations of the RNAi-mutant Avans-1/18

Plant No., generation	Number of kernels, endosperm type				Ratio (modified: regular)	χ^2	p
	Total	Modified		With a regular vitreous layer			
		floury	with vitreous inclusions				
No. 1 T_0 <i>ubi1+bar+</i>	18	13	5 (around the periphery)	–	–	–	–
1-3, T_1 <i>ubi1+bar+</i>	7	7	–	–	–	–	–
1-6, T_1 <i>ubi1-bar-</i>	8	–	–	8	–	–	–
190-1, T_1 <i>ubi1+bar+</i>	72	58	–	14	3:1	1.185	0.50–0.25
190-2, T_1 <i>ubi1+bar+</i>	87	58	16	13	3:1	4.693	< 0.05
190-3 T_1 <i>ubi1-bar+</i>	16	–	–	16	–	–	–
190-4, T_1 <i>ubi1+bar+</i>	29	29	–	–	–	–	–
1-3-5, T_2 (from T_1 1-3) <i>ubi1+bar+</i>	8	8	–	–	–	–	–
1-3-10, T_2 (from T_1 1-3) <i>ubi1+bar+</i>	8	4	4 (around the periphery)	–	–	–	–
181-II-5, T_2 (from T_1 190-1) <i>ubi1+bar+</i>	353	99	152	102	3:1	2.856	0.05–0.01
182-I-8, T_2 (from T_1 190-2) <i>ubi1+bar+</i>	70	55	11	4	15:1	0.034	0.90–0.75
182-II-3, T_2 (from T_1 190-2) <i>ubi1+bar+</i>	288	114	149	24	15:1	2.133	0.25–0.10
182-I-10, T_2 (from T_1 190-2) <i>ubi1+bar+</i>	267	149	118	–	Homozygote		
184-I-8, T_2 (from T_2 190-4) <i>ubi1+bar+</i>	407	148	229	–	Homozygote		
184-I-2, T_2 (from T_2 190-4) <i>ubi1+bar+</i>	70	52	18	–	Homozygote		
150a-1, T_3 (from T_2 182-I-8) <i>ubi1+bar+</i>	93	26	17+50 (around the periphery)	7	15:1	0.096	0.90–0.75
150a-12, T_3 (from T_2 182-I-8) <i>ubi1+bar+</i>	100	16	26+58 (around the periphery)	–	Homozygote		
150a-15, T_3 (from T_2 182-I-8) <i>ubi1+bar+</i>	100	26	52+22 (around the periphery)	–	Homozygote		
152a-10, T_3 (from T_2 184-I-8) <i>ubi1+bar+</i>	101	33	19+49 (around the periphery)	–	Homozygote		
197-6, T_3 (from T_2 1-3-5) <i>ubi1+bar+</i>	311	166	116+29 (around the periphery)	–	Homozygote		
153a-6, T_4 (from T_3 197-6) <i>ubi1+bar+</i>	100	33	9+58 (around the periphery)	–	Homozygote		
153a-4, T_4 (from T_3 197-6) <i>ubi1+bar+</i>	100	35	13+52 (around the periphery)	–	Homozygote		

Wang et al., 2021), was used as a reference control. In addition, the copy number of the introduced genetic construct was also calculated using DNA from the plant No. 124-3-9 (T_4), which was segregated as a mono-heterozygote (3:1), was also used as a reference control (see Table 2).

It was found that in those plants that, based on the results of segregation analysis for the *bar* gene, were expected to have one copy of the genetic construct, the copy number, according to qPCR data, was the lowest (0.6...0.9; on average 0.76 ± 0.06) (Table 4). At the same time, in the plant No. 124-3-3 – a putative homozygote with two copies of the genetic construct – the copy number, according to qPCR data, was 2 times higher (1.4 ± 0.2). These data show that our qPCR analysis quite accurately reflects the copy number of the construct under study, and the results obtained using it are trustworthy.

The analysis revealed variations in the number of copies of the genetic construct for RNA silencing in different T_3 plants. Thus, plant No. 150a-15 from the T_1 190-2 family had two copies of the construct (see Table 4). Considering the lack of segregation for the *bar* gene (see Table 2) and for the endosperm type in the kernels formed on its panicle (see Table 3), it is obvious that it is a mono-homozygote. At the same time, in the plant No. 150a-12, the copy number was significantly higher, indicating the presence of three-four copies of the genetic construct. The presence of individuals with two and four copies of the construct from the progeny of the same T_1 plant (190-2) indicates the presence of two independent T-DNA insertions. This assumption is confirmed by the di-hybrid segregation pattern (15:1) for the *bar* gene and for the endosperm type in the progeny of some plants from the 190-2 family.

Thus, the data of classical genetics and qPCR quite accurately agree with each other and indicate the presence of at least two independent copies of the genetic construct for RNA silencing in the parental T_0 transgenic plant. Noteworthy is the high copy number of the construct in plants from the progenies of T_1 190-4 and 1-3 plants in T_3 and T_4 generations, which carried at least four copies of the genetic construct.

In vitro digestibility of grain proteins. An assessment of the digestibility of grain proteins obtained from whole-ground kernels set on panicles of T_1 and T_2 plants showed the inheritance of a high level of protein digestibility found in T_0 generation, although the values in most T_2 plants were slightly lower than the values of this trait in T_1 . Plants grown with artificial watering during grain maturation (2 times per 7–8 l/m²) had higher digestibility compared to plants grown in the absence of additional watering (Table 5). Apparently, conditions of higher water availability contribute to more efficient expression of the genetic construct for silencing. Kernels with a floury endosperm type and with endosperm containing inclusions of vitreous endosperm had the same level of protein digestibility (Fig. 3). At the same time, kernels with a regular vitreous endosperm had lower digestibility compared to kernels with a mutant type of endosperm (floury or with vitreous inclusions).

Remarkably, in the plant 184-I-7, an unusual floury type endosperm with a pink coloration was found. This coloration is possibly conditioned by polyphenols, which normally determine the color of the vitreous layer of the endosperm in the kernels of the original cv. Avans. The flour obtained from

Table 4. The number of copies of the genetic construct for RNA silencing of the *gKAF1* gene in different plants from the T_3 and T_4 generations according to qPCR results

Transgenic plant	Copy number in relation	
	to <i>APRT</i>	to <i>nos</i> -promoter ¹
T_1 190-2 progeny		
124-3-9 T_4 , mono-heterozygote	0.8 ± 0.0	–
124-3-10 T_4 , mono-heterozygote	0.9 ± 0.2	1.1 ± 0.3
124-3-4 T_4 , mono-heterozygote	0.8 ± 0.1	1.0 ± 0.2
124-3-8 T_4 , mono-heterozygote	0.6 ± 0.0	0.8 ± 0.1
On average, for mono-heterozygotes	0.76 ± 0.06	
T_1 190-4 progeny		
156-2 T_2	4.0 ± 0.5	5.1 ± 0.7
152a-3 T_3	4.1 ± 0.2	5.1 ± 0.4
152a-10 T_3	4.2 ± 0.3	5.3 ± 0.5
152a-14 T_3	3.3 ± 0.3	4.1 ± 0.5
T_1 1-3 progeny		
153a-4 T_4	4.0 ± 0.2	5.0 ± 0.4
153a-6 T_4	2.9 ± 0.2	3.7 ± 0.3
153a-3 T_4	3.3 ± 0.3	4.1 ± 0.4
153b-2 T_4	3.5 ± 0.7	4.4 ± 0.9
153b-1 T_4	2.0 ± 0.1	2.6 ± 0.2
159-5 non-transgenic sibling	0.02 ± 0.0	0.03 ± 0.0
Avans, original cv.	0.03 ± 0.0	0.04 ± 0.0

¹ Copy number was calculated in relation to mono-heterozygote, No. 124-3-9 (T_4).

the same plant had a similar coloration. The digestibility of proteins in such flour was slightly lower than that of flour from other plants, possibly due to the inhibition of pepsin action by polyphenols.

The total protein content in the grain of T_1 plants decreased by 7.5–17.3 % compared to the original cv. Avans: 12.8–14.3 % vs. 15.5 %, while the percentage of lysine in the grains of plant No. 190-4 increased by 75 % (see Supplementary Material)¹, which may have been a consequence of re-balancing protein synthesis in the grains of transgenic plants and the appearance of proteins with higher lysine content

¹ Supplementary Material is available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Elkonin_Engl_28_1.pdf

Table 5. *In vitro* grain protein digestibility in plants from the progeny of the Avans-1/18 mutant carrying the genetic construct for RNA silencing of the *gKAF1* gene

Plant	Additional watering ¹	Endosperm type	Digestibility, %	Grain protein content, %
T ₁				
190-1 <i>ubi1+bar+</i>	+	Floury	–	12.8 a
190-3 <i>ubi1–bar+</i>	+	Regular vitreous	75.3 b	14.5 b
190-4 <i>ubi1+bar+</i>	+	Floury	92.2 c	14.3 b
Avans, original cv.	+	Regular vitreous	52.3 a	15.5 c
<i>F</i>			558.946*	51.97*
LSD ₀₅			5.1	0.69
T ₂ (from T ₁ 190-4)				
184-I-8	+	Floury	79.8 h	9.16 a
		With vitreous inclusions	81.0 hi	
184-I-7	+	Floury, pink	62.4 a	–
184-I-2	+	Floury	75.0 efg	11.13 de
116-4	–		71.0 bcde	–
116-5	–		72.2 def	9.36 a
182-I-8	+		76.1 efg	11.63 ef
182-II-3	+	Floury	79.3 h	–
		With vitreous inclusions	77.6 gh	–
		Regular vitreous	71.4 cdef	–
T ₂ (from T ₁ 190-2)				
124-3	–	Floury	63.6 a	13.22 h
		Regular vitreous	64.4 a	
T ₃ (from T ₁ 1-3)				
197-5	+	Floury	80.3 h	11.88 f
197-6	+		78.8 gh	12.78 gh
Avans, original cv.	+	Regular vitreous	59.3 a	11.09 de
<i>F</i>			16.988**	37.257*
LSD ₀₅			5.735	0.618

¹ Additional watering: 2 times per 8 l/m² during the period of grain maturation; * $p < 0.05$, ** $p < 0.01$; data denoted by different letters significantly differ at $p < 0.05$ according to Duncan's Multiple Range Test.

compared to γ -kafirin. This re-balancing of protein content has been previously described in many cereal RNAi mutants with genetic constructs that suppress prolamine synthesis (Elkonin et al., 2016a). The protein content in the grain of different T₂ plants varied from 9.2 to 12.8 % and in most of the studied plants it did not differ significantly from the original cv. Avans (11.1 %). In two plants from family No. 190-4 (184-I-8 and 116-5), apparently containing two copies of the genetic construct, the protein content was significantly lower (9.2–9.4 %). However, the influence of the number of copies of the genetic construct for silencing the *gKAF1* gene on the total protein content in grain requires additional investigation.

Analysis of agronomically valuable traits. To study the manifestation of agronomically important traits in the plants carrying the genetic construct for RNA silencing of the γ -kafirin gene, in 2020, a few T₁ plants were grown in an experimental plot under field conditions. It was found that these plants have a reduced height compared to the original cv. Avans: 82.3 ± 1.7 cm and 100.8 ± 3.4 cm, respectively ($p < 0.01$). These differences were also observed in T₂ generation: 81.5–91.6 cm, on average, in different T₂ families, compared to 104.1 cm for the original cultivar ($p < 0.05$) (2021 data). In this connection, for a more detailed study of the manifestation of the main agronomically valuable traits in the mutant, T₃ families (the progeny of three T₁ plants Nos. 190-2,

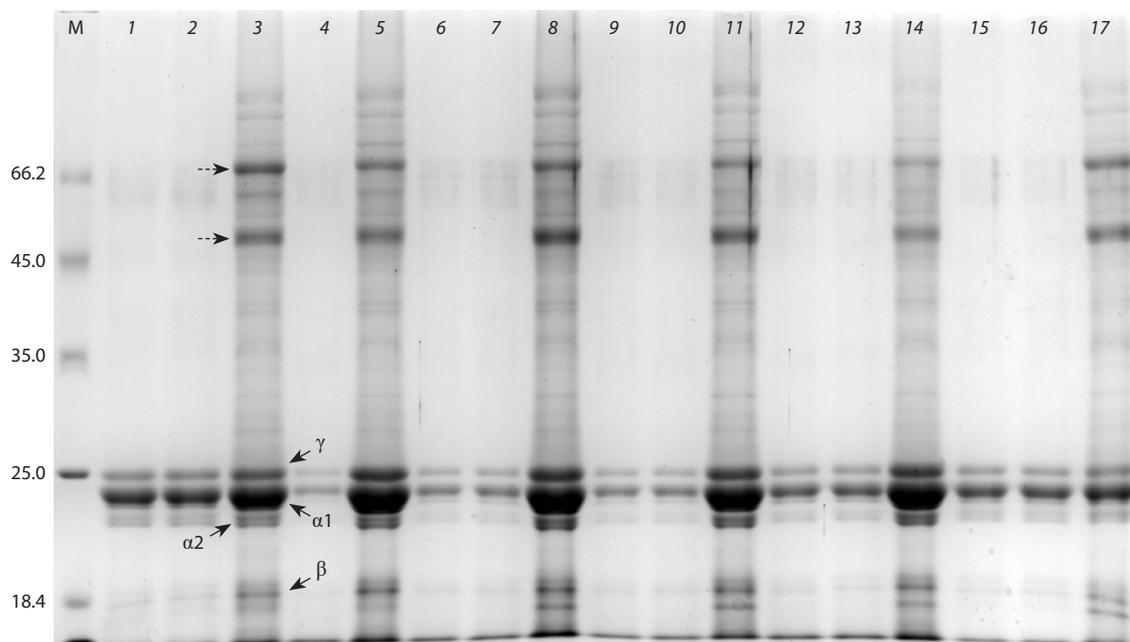


Fig. 3. SDS-PAGE of grain proteins of the original cv. Avans and plants of Avans-1/18 after pepsin treatment of the flour. 1–3, 15–17 – original cv. Avans; 4, 5 – T₁ 190-4; 6–11 – T₂ 184-l-8 (progeny from 190-4); 6–8 – grains with floury endosperm; 9–11 – grains with floury endosperm interspersed with vitreous endosperm; 12–14 – T₂ 184-l-7 (progeny from 190-4) – grains with pink floury endosperm. 1, 2, 4, 6, 7, 9, 10, 12, 13, 15, 16 – after treating of the flour with pepsin; 3, 5, 8, 11, 14, 17 – without pepsin treatment (control). The positions of different classes of kafirins are marked with arrows. Dotted arrows indicate proteins belonging to the globulin fraction.

Table 6. Manifestation of some agronomically important traits in the progeny of the mutant Avans-1/18 (averages from three replications)

Line	Plant height, cm	Peduncle length, cm	1000 grain weight, g	Grain yield per panicle, g
Avans, original cv.	104.6 c	12.0 b	32.5 c	32.2 c
T ₃ 150 (from T ₁ 190-2) <i>ubi1</i> + ¹	91.6 ab	7.4 a	29.7 abc	24.6 a
T ₃ 151 (from T ₁ 190-3) <i>ubi1</i> -	91.0 ab	7.4 a	31.8 bc	27.1 ab
T ₃ 152 (from T ₁ 190-4) <i>ubi1</i> + ¹	86.6 a	5.1 a	28.0 a	22.2 a
T ₄ 153 (from T ₁ 1-3) <i>ubi1</i> + ¹	97.9 bc	7.4 a	29.1 ab	26.0 b
<i>F</i>	10.764*	10.848*	3.446*	12.051*
LSD ₀₅	7.0	2.5	3.0	2.9

¹ The presence of the genetic construct for RNA silencing of the *gKAF1* gene (based on the results of PCR with primers for the *ubi1*-intron). Data denoted by different letters are significantly different at $p < 0.05$, according to Duncan's Multiple Range Test.

190-3, 190-4, which were grown in the field), and one T₄ family (the progeny of T₁ plant No. 1-3 from the greenhouse) were grown in 2022 in experimental plots, in 3 replications.

It was found that all the studied lines of transgenic origin derived from a single initial T₀ plant differ from the original cv. Avans in reduced plant height, shortened peduncle and reduced grain yield per panicle (Table 6). Such a decrease in the manifestation of these traits in mutants carrying genetic constructs for RNA silencing of prolamin genes is described for the first time in cereals. Such a change in plant height and peduncle length may be a consequence of off-target effects of expression of a genetic construct for RNA silencing, which are well known and described in the literature (Jackson et al.,

2003; Senthil-Kumar, Mysore, 2011). However, this explanation seems unlikely since in the line No. 151 (progeny of plant 190-3), this construct was eliminated, while the short stature and shortened peduncle were preserved.

Another explanation may be the insertion of the genetic construct for RNA silencing into one of the loci that controls plant height and peduncle length. Similar facts of induction of mutations in genes encoding various plant traits as a result of T-DNA insertion are widely known (Wilson et al., 2006; Deineko et al., 2007; Ram et al., 2019). In most cases, such insertions lead to plant DNA deletions that disrupt the functional activity of genes. Previously, mutations of short stature caused by T-DNA insertion, including mutation of the locus

that controls gibberellin synthesis, which is involved in the regulation of plant height, were described in Arabidopsis (Feldmann et al., 1989; Chiang et al., 1995). In sorghum, we were unable to identify any reports on the induction of mutants by T-DNA insertion in the available literature.

The weight of 1000 grains was significantly reduced in two lines carrying the RNA-silencing construct (Nos. 152, 153). Obviously, such a decrease in the weight of 1000 grains is a consequence of modification of the endosperm type, since in the loose flourey endosperm, there are air cavities between the starch granules. It is noteworthy that in line No. 151 (progeny of plant 190-3), in which the kernels, due to the elimination of the RNA-silencing construct, had a regular endosperm with a thick vitreous layer, there were no significant differences for this trait from the original cv. Avans. The absence of significant differences between the original cultivar and line No. 150 is also possibly explained by the presence of heterozygous plants, the panicles of which contain kernels with vitreous endosperm (see Table 3). Thus, the decrease in 1000-grain weight is a direct consequence of the expression of the *gKAF1* gene RNA-silencing construct. A similar decrease in 1000-grain weight was described in sorghum RNAi mutants obtained in the African variety P898012, carrying a genetic construct for RNA silencing of two γ -kafirin genes, 27 and 50 kDa (Ndimba et al., 2017).

Such a decrease in grain weight could not but cause a decrease in grain yield per panicle, which was observed in all transgenic lines we obtained, and in this regard, a change in this trait is also a consequence of the expression of the introduced genetic construct. At the same time, the drop in grain yield may also be due to a reduced development power of plants due to T-DNA insertion into one of the loci that controls plant height, since this trait was also reduced in line No. 151, in which the silencing construct was eliminated. In this regard, a decrease in grain yield per panicle may also be a consequence of insertion mutagenesis induced by T-DNA.

It should be noted that the presence of several unlinked copies of the genetic construct, apparently integrated into different regions of the genome, which, due to the effect of gene position, may differ in their effect on the phenotypic traits of plants, opens up the possibility of selecting lines that combine improved protein digestibility with the required manifestation of agronomically valuable traits. Thus, a line was identified (No. 153, T₄ derived from T₁ 1-3), in which plant height and grain yield per panicle were reduced to a lesser extent (by 5–6 %), while the level of protein digestibility reached 81 % (see Table 5). In this regard, this line may be of greater interest for practical breeding.

Conclusion

The results of this study indicate that the genetic construct for RNA silencing of the sorghum γ -kafirin gene (*gKAF1*) is represented in the genome of different plants from T₃ and T₄ generations of the RNAi mutant Avans-1/18 in 2–4 copies, which, apparently, are a consequence of two independent events of T-DNA integration. This construct is stably inherited and expressed in the T₃ and T₄ generations, modifying the endosperm texture, improving protein digestibility (up to 81 %, compared to 52–59 % in the original cultivar), and increasing the percentage of lysine (by 75 %). At the same time, cases

of disturbances of the integrity of the integrated T-DNA and elimination of the genetic construct for silencing have been recorded. Transgenic plants of the RNAi mutant Avans-1/18 are characterized by reduced plant height, shortened peduncle, reduced 1000-grain weight, and panicle grain yield. A line (T₃) with high digestibility of endosperm proteins (81 %) and a minimal decrease in agronomically valuable traits (by 5–7 %) was isolated.

References

- Aboubacar A., Axtell J.D., Huang C.P., Hamaker B.R. A rapid protein digestibility assay for identifying highly digestible sorghum lines. *Cereal Chem.* 2001;78(2):160-165. DOI 10.1094/CCHEM.2001.78.2.160
- Bharathi J.K., Anandan R., Benjamin L.K., Muneer S., Prakash M.A.S. Recent trends and advances of RNA interference (RNAi) to improve agricultural crops and enhance their resilience to biotic and abiotic stresses. *Plant Physiol. Biochem.* 2023;194:600-618. DOI 10.1016/j.plaphy.2022.11.035
- Borisenko N., Elkonin L., Kenzhegulov O. Inheritance of the genetic construct for RNA-silencing of the γ -kafirin gene (*gKAF1*) in the progeny of transgenic sorghum plants. *BIO Web Conf.* 2022;43:03015. DOI 10.1051/bioconf/20224303015
- Casu R.E., Selivanova A., Perroux J.M. High-throughput assessment of transgene copy number in sugarcane using real-time quantitative PCR. *Plant Cell Rep.* 2012;31(1):167-177. DOI 10.1007/s00299-011-1150-7
- Chiang H.H., Hwang I., Goodman H.M. Isolation of the *Arabidopsis G44* locus. *Plant Cell.* 1995;7(2):195-201. DOI 10.1105/tpc.7.2.195
- da Silva L.S. Transgenic sorghum: Effects of altered kafirin synthesis on kafirin polymerisation, protein quality, protein body structure and endosperm texture. Thesis. Pretoria: University of Pretoria South Africa, 2012
- da Silva L.S., Taylor J., Taylor J.R.N. Transgenic sorghum with altered kafirin synthesis: kafirin solubility, polymerization, and protein digestion. *J. Agric. Food Chem.* 2011a;59(17):9265-9270. DOI 10.1021/jf201878p
- da Silva L.S., Jung R., Zhao Z., Glassman K., Grootboom A.W., Mehlo L., O'Kennedy M.M., Taylor J., Taylor J.R.N. Effect of suppressing the synthesis of different kafirin subclasses on grain endosperm texture, protein body structure and protein nutritional quality in improved sorghum lines. *J. Cereal Sci.* 2011b;54(1):160-167. DOI 10.1016/j.jcs.2011.04.009
- Deineko E.V., Zagorskaya A.A., Shumny V.K. T-DNA-induced mutations in transgenic plants. *Russ. J. Genet.* 2007;43(1):1-11. DOI 10.1134/S1022795407010012
- de Mesa-Stonestreet N.J., Alavi S., Bean S.R. Sorghum proteins: the concentration, isolation, modification, and food applications of kafirins. *J. Food Sci.* 2010;75(5):90-104. DOI 10.1111/j.1750-3841.2010.01623.x
- Duressa D., Weerasoriya D., Bean S.R., Tilley M., Tesso T. Genetic basis of protein digestibility in grain sorghum. *Crop Sci.* 2018;58(6):2183-2199. DOI 10.2135/cropsci2018.01.0038
- Elkonin L.A., Italianskaya J.V., Fadeeva I.Yu., Bychkova V.V., Kozhemyakin V.V. *In vitro* protein digestibility in grain sorghum: effect of genotype and interaction with starch digestibility. *Euphytica.* 2013;193:327-337. DOI 10.1007/s10681-013-0920-4
- Elkonin L.A., Domanina I.V., Ital'yanskaya Yu.V. Genetic engineering as a tool for modification of seed storage proteins and improvement of nutritional value of cereal grain. *Agric. Biol.* 2016a;51(1):17-30. DOI 10.15389/agrobiol.2016.1.17eng
- Elkonin L.A., Italianskaya J.V., Domanina I.V., Selivanov N.Y., Rakin A.L., Ravin N.V. Transgenic sorghum with improved digestibility of storage proteins obtained by *Agrobacterium*-mediated transformation. *Russ. J. Plant Physiol.* 2016b;63(5):678-689. DOI 10.1134/S1021443716050046

- Elkonin L.A., Panin V.M., Kenzhegulov O.A., Sarsenova S.Kh. RNAi-mutants of *Sorghum bicolor* (L.) Moench with improved digestibility of seed storage proteins. In: Jimenez-Lopez J.C. (Ed.). Grain and Seed Proteins Functionality. London: Intech Open Ltd., 2021;1-17. DOI 10.5772/intechopen.96204
- Feldmann K.A., Marks M.D., Christianson M.L., Quatrano R.S. A Dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science*. 1989;243(4896):1351-1354. DOI 10.1126/science.243.4896.1351
- Grootboom A.W., Mkhonza N.L., Mbambo Z., O'Kennedy M.M., da Silva L.S., Taylor J., Taylor J.R.N., Chikwamba R., Mehlo L. Co-suppression of synthesis of major α -kafirin sub-class together with γ -kafirin-1 and γ -kafirin-2 required for substantially improved protein digestibility in transgenic sorghum. *Plant Cell Rep.* 2014; 33(3):521-537. DOI 10.1007/s00299-013-1556-5
- Guo Q., Liu Q., Smith N.A., Liang G., Wang M.B. RNA silencing in plants: mechanisms, technologies and applications in horticultural crops. *Curr. Genomics*. 2016;17(6):476-489. DOI 10.2174/1389202917666160520103117
- Jackson A.L., Bartz S.R., Schelzer J., Kobayashi S.V., Burchard J., Mao M., Li B., Cavet G., Linsley P.S. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* 2003;21(6): 635-638. DOI 10.1038/nbt831
- Kumar T., Dweikat I., Sato S., Ge Z., Nersesian N., Chen H., Elthon T., Bean S., Ioerger B.P., Tilley M., Clemente T. Modulation of kernel storage proteins in grain sorghum (*Sorghum bicolor* (L.) Moench). *Plant Biotechnol. J.* 2012;10:533-544. DOI 10.1111/j.1467-7652.2012.00685.x
- Muhammad T., Zhang F., Zhang Y., Liang Y. RNA interference: a natural immune system of plants to counteract biotic stressors. *Cells*. 2019;8(1):38. DOI 10.3390/cells8010038
- Ndimba R.J., Kruger J., Mehlo L., Barnabas A., Kossmann J., Ndimba B.K. A comparative study of selected physical and biochemical traits of wild-type and transgenic sorghum to reveal differences relevant to grain quality. *Front. Plant Sci.* 2017;8:952. DOI 10.3389/fpls.2017.00952
- Nunes A., Correia I., Barros A., Delgado I. Sequential *in vitro* pepsin digestion of uncooked and cooked sorghum and maize samples. *J. Agric. Food Chem.* 2004;52(7):2052-2058. DOI 10.1021/jf0348830
- Oria M.P., Hamaker B.R., Axtell J.D., Huang C.P. A highly digestible sorghum mutant cultivar exhibits a unique folded structure of endosperm protein bodies. *Proc. Natl. Acad. Sci. USA.* 2000;97(10): 5065-5070. DOI 10.1073/pnas.080076297
- Ram H., Soni P., Salvi P., Gandass N., Sharma A., Kaur A., Sharma T.R. Insertional mutagenesis approaches and their use in rice for functional genomics. *Plants*. 2019;8(9):310. DOI 10.3390/plants8090310
- Senthil-Kumar M., Mysore K.S. Caveat of RNAi in plants: the off-target effect. *Methods Mol. Biol.* 2011;744:13-25. DOI 10.1007/978-1-61779-123-9_2
- Smith N.A., Singh S.P., Wang M.-B., Stoutjesdijk P., Green A., Waterhouse P.M. Total silencing by intron-spliced hairpin RNAs. *Nature*. 2000;407(6802):319-320. DOI 10.1038/35030305
- Tesso T., Ejeta G., Chandrashekar A., Huang C.P., Tandjung A., Lewamy M., Axtell J., Hamaker B.R. A novel modified endosperm texture in a mutant high-protein digestibility/high-lysine grain sorghum (*Sorghum bicolor* (L.) Moench). *Cereal Chem.* 2006;83(2):194-201. DOI 10.1094/CC-83-0194
- Wang L., Gao L., Liu G., Meng R., Liu Y., Li J. An efficient sorghum transformation system using embryogenic calli derived from mature seeds. *PeerJ*. 2021;9:e11849. DOI 10.7717/peerj.11849
- Wilson A.K., Latham J.R., Steinbrecher R.A. Transformation-induced mutations in transgenic plants: Analysis and biosafety implications. *Biotechnol. Genet. Eng. Rev.* 2006;23:209-238. DOI 10.1080/02648725.2006.10648085
- Wong J.H., Lau T., Cai N., Singh J., Pedersen J.F., Vensel W.H., Hurkman W.J., Wilson J.D., Lemaux P.G., Buchanan B.B. Digestibility of protein and starch from sorghum (*Sorghum bicolor*) is linked to biochemical and structural features of grain endosperm. *J. Cereal Sci.* 2009;49(1):73-82. DOI 10.1016/j.jcs.2008.07.013
- Zhuravlyov V.S., Dolgikh V.V., Timofeev S.A., Gannibal F.B. RNA interference method in plant protection against insect pests. *Vestnik Zashchity Rastenij = Plant Protection News*. 2022;105(1):28-39. DOI 10.31993/2308-6459-2022-105-1-15219 (in Russian)

ORCID

L.A. Elkonin orcid.org/0000-0003-3806-5697
N.V. Borisenko orcid.org/0000-0003-3543-9083
T.E. Pylaev orcid.org/0000-0002-2701-3333

O.A. Kenzhegulov orcid.org/0000-0002-1272-3105
S.Kh. Sarsenova orcid.org/0009-0007-2471-1631
N.Yu. Selivanov orcid.org/0000-0002-7438-0013
V.M. Panin orcid.org/0009-0007-2851-415X

Acknowledgements. The work was carried out with financial support from the budget theme of the Federal Center of Agriculture Research of the South-East Region (Saratov, Russia) FNWF-2023-0006 and the state task of the Ministry of Science and Higher Education of the Russian Federation for the Federal Research Center "Saratov Scientific Center of the Russian Academy of Sciences", theme No. 121031700141-7. HPLC analysis of the amino acid composition of flour proteins was carried out using the equipment of the Center for Collective Use "Symbiosis" of the Institute of Biochemistry and Physiology of Plants and Microorganisms of RAS (Saratov, Russia). qPCR analysis of the copy number of the genetic construct was carried out at the Research and Education Center for Molecular Genetic and Cellular Technologies of the V.I. Razumovsky Saratov State Medical University.

Conflict of interest. The authors declare no conflict of interest.

Received August 16, 2023. Revised September 25, 2023. Accepted October 13, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Halo-RPD: searching for RNA-binding protein targets in plants

A.O. Shamustakimova

All-Russian Research Institute of Agricultural Biotechnology, Moscow, Russia

✉ nastja_sham@mail.ru

Abstract. Study of RNA-protein interactions and identification of RNA targets are among the key aspects of understanding RNA biology. Currently, various methods are available to investigate these interactions with, RNA immunoprecipitation (RIP) being the most common. The search for RNA targets has largely been conducted using antibodies to an endogenous protein or to GFP-tag directly. Having to be dependent on the expression level of the target protein and having to spend time selecting highly specific antibodies make immunoprecipitation complicated. Expression of the GFP-fused protein can lead to cytotoxicity and, consequently, to improper recognition or degradation of the chimeric protein. Over the past few years, multifunctional tags have been developed. SNAP-tag and HaloTag allow the target protein to be studied from different perspectives. Labeling of the fusion protein with custom-made fluorescent dyes makes it possible to study protein expression and to localize it in the cell or the whole organism. A high-affinity substrate has been created to allow covalent binding by chimeric proteins, minimizing protein loss during protein isolation. In this paper, a HaloTag-based method, which we called Halo-RPD (HaloTag RNA PullDown), is presented. The proposed protocol uses plants with stable fusion protein expression and Magne® HaloTag® magnetic beads to capture RNA-protein complexes directly from the cytoplasmic lysate of transgenic *Arabidopsis thaliana* plants. The key stages described in the paper are as follows: (1) preparation of the magnetic beads; (2) tissue homogenization and collection of control samples; (3) precipitation and wash of RNA-protein complexes; (4) evaluation of protein binding efficiency; (5) RNA isolation; (6) analysis of the RNA obtained. Recommendations for better NGS assay designs are provided.

Key words: *A. thaliana*; HaloTag; RNA-binding proteins; RNA pulldown assay; RNA-protein complexes; cold-shock domain protein.

For citation: Shamustakimova A.O. Halo-RPD: searching for RNA-binding protein targets in plants. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(1):74-79. DOI 10.18699/vjgb-24-09

Нало-RPD: в поисках мишеней РНК-связывающих белков растений

А.О. Шамустакимова

Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва, Россия

✉ nastja_sham@mail.ru

Аннотация. Изучение РНК-белковых взаимодействий и идентификация РНК-мишеней относятся к важнейшим аспектам понимания биологии РНК. К настоящему времени предложены различные методы изучения таких взаимодействий; одним из широко распространенных является иммунопреципитация РНК (RIP). Большинство работ по поиску РНК-мишеней было проведено с использованием антител непосредственно на эндогенный белок или же на GFP, слитый с целевым белком. Зависимость от уровня экспрессии целевого белка и подбор специфичных антител значительно затрудняют классическую иммунопреципитацию. Белок же, слитый с GFP, нередко может быть цитотоксичен, что в дальнейшем приведет к его неправильному узнаванию и/или деградации. В последние годы был разработан ряд мультифункциональных тагов, включая SNAP-tag и HaloTag. Такие таги способствуют изучению целевых белков с разных сторон. Для них созданы флуоресцентные красители, способные прочно связываться с определенным участком тага. Это позволяет как изучать наработку химерного белка, так и определять его локализацию непосредственно в клетке или во всем организме. Для таких тагов разработаны также высокоаффинные субстраты, ковалентно связывающие химерные белки, что значительно сокращает потери в ходе выделения. В данной работе представлен метод, основанный на системе HaloTag, который мы назвали Halo-RPD (HaloTag RNA PullDown). В протоколе используются растения со стабильной экспрессией химерного белка и магнитные шарики Magne® HaloTag® Beads для захвата РНК-белковых комплексов непосредственно из цитоплазматического лизата трансгенных растений *Arabidopsis thaliana*. Приводится описание основных этапов: 1) подготовка магнитных шариков; 2) гомогенизация тканей и отбор контролей; 3) осаждение и отмывка РНК-белковых комплексов; 4) определение эффективности связывания белка; 5) выделение РНК; 6) анализ полученной РНК. Даны рекомендации для планирования эксперимента по высокопроизводительному секвенированию. Ключевые слова: *A. thaliana*; HaloTag; РНК-связывающие белки; соосаждение; РНК; РНК-белковые комплексы; белок с доменом холодового шока.

Introduction

RNA-binding proteins play a major part in complex cellular processes, such as differentiation, development, responses to biotic and abiotic stress factors, and post-transcriptional control.

In recent years, the variety of methods for studying RNA-protein interactions has expanded significantly (Ramanathan et al., 2019). However, some older technologies, such as RNA-immunoprecipitation (RIP), still remain rather common (Brooks, Rigby, 2000). The latter is based on *in vivo* mapping of RNA-protein interactions using crosslinking agents, such as ultraviolet or formaldehyde. Currently, RIP is used in the vast majority of studies investigating RNA-protein complexes in plants (Köster, Meyer, 2018; Frydrych Capelari et al., 2019; Seo, Chua, 2019; Steffen et al., 2019).

Despite its wide use, RIP has several downsides, e.g. UV radiation induces the formation of the irreversible covalent bond between a protein and RNA; formaldehyde not only binds the protein of interest to RNA but crosslinks its partner proteins as well; highly specific antibodies are required for a successful outcome.

HaloTag-fused proteins are used in the proposed protocol as an alternative to RIP (Los et al., 2008). Initially, the HaloTag technology was intended and successfully used for precipitation of protein-protein and DNA-protein complexes from bacterial and mammalian cell lysate (Urh et al., 2008). However, in the last few years, the technology was adapted (van Dijk et al., 2015; Banks et al., 2016; Li et al., 2020) and modified (Gu et al., 2018) to identify RNA-protein complexes in tissue cells in humans and animals.

So far, there have been only two papers, the authors of which investigated the use of HaloTag technology for the search of partner proteins in plants (Samanta, Thakur, 2017; Ren et al., 2020). In the first paper, the authors analyzed mediator proteins in transgenic plants of rice, and in the second one, the binding site of transcription factor *ZmNST3* in transgenic plants of maize was investigated.

The goal of the present study was to design a protocol for isolating RNA-protein complexes from the cytoplasm of *Arabidopsis thaliana* plants using the HaloTag technology. Papers (Sorenson, Bailey-Serres, 2015; Banks et al., 2016) were used as a reference to design the protocol called **HaloTag RNA-PullDown** (Halo-RPD).

The Halo-RPD method

Plant material. Two lines of *Arabidopsis thaliana* (L.) Heynh. Columbia ecotype were used in the study as transgenic plants with stable expression of HaloTag and *EsCSDP3*-HaloTag-fused proteins (Taranov et al., 2018).

The stages of HaloRPD protocol for isolation of RNA-protein complexes are as follows: preparation of magnetic beads; tissue homogenization, and collection of control samples; pull-down and wash of RNA-protein complexes; evaluation of protein binding efficiency; RNA isolation; analysis of the obtained RNA. A simplified assay design is presented in Figure 1.

Preparation of magnetic beads

The Magne HaloTag Beads (Promega Corp.) magnetic particle suspension (100 µl) was placed in two tubes for the experimental (*EsCSDP3*-HaloTag) and control (HaloTag) samples.

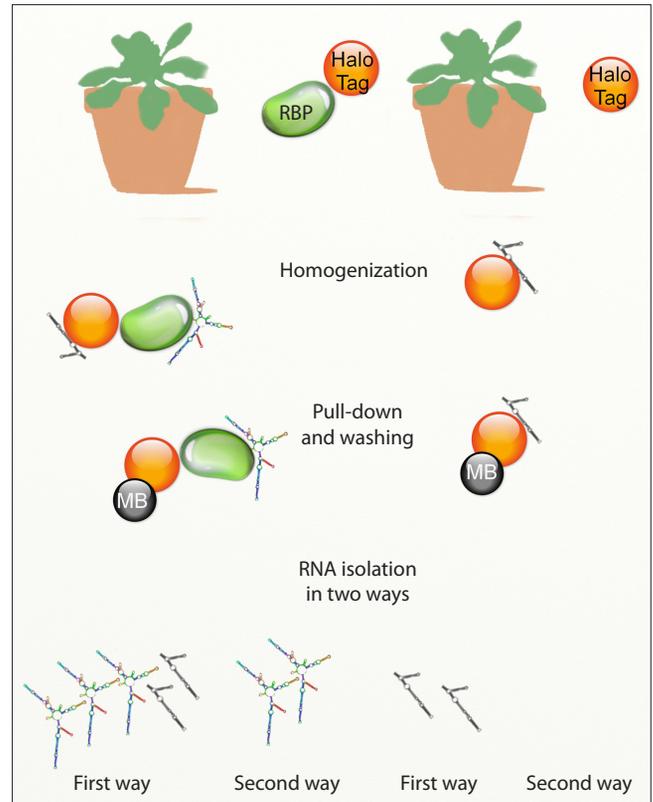


Fig. 1. A simplified Halo-RPD assay design. Transgenic plants expressing HaloTag-fused RNA-binding protein (on the left) and plants expressing the HaloTag protein (on the right).

Plant tissue is homogenized to obtain cytoplasmic lysate and incubated with the magnetic beads capable of binding the HaloTag protein by covalent bonds. The beads with precipitated target protein complexes are washed, and RNA is isolated in two ways. The first way is to isolate RNA directly from the magnetic beads by incubation in the ExtractRNA reagent, and the second implies elution in TEV buffer and further isolation using the ExtractRNA reagent.

The tubes were incubated on a magnetic stand until transparent, and the liquid was carefully removed without disturbing the beads. The buffer (400 µl) composed of 50 mM Tris-HCl (pH 7.4), 137 mM NaCl, 2.7 mM KCl, and 0.05 % Igepal Ca-630 (Promega Corp.) was added to the beads, and the suspension was gently mixed manually several times. The tubes were transferred to a magnetic stand, incubated until transparent, and liquid was fully removed by pipetting. The tubes were then washed two more times. The supernatant was not removed after the third wash. The tubes were stored at +4 °C.

The manufacturer offers two types of substrate for pull-down of fusion protein from lysate, namely HaloLink resin (Promega Corp.) and MagneBeads (Promega Corp.), the latter being a newer product. The magnetic beads have an advantage of high binding affinity of HaloTag-fused proteins and low non-specific binding level. For instance, 1 ml of magnetic beads binds over 20 mg of protein, whereas the same amount of resin only binds 7 mg. The resin was used in prior assays of our study, which significantly increased their duration due to multiple centrifugation steps. This approach also required the availability of LowBind tubes or silicone treatment of the tubes to minimize resin loss.

Tissue homogenization and collection of control samples

For better preservation of plant tissue, leaf blades were wrapped in foil and placed in liquid nitrogen.

Precooled mortars and pestles were used for homogenization. The tissue was placed in a mortar, a small amount of liquid nitrogen was added, and the mixture was homogenized to a powder. The obtained powder was then added to a tube with 300 μ l of lysis buffer composed of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 % Triton X-100, 0.1 mM benzamidine HCl, 55 μ M phenanthroline, 10 μ M bestatin, 20 μ M leupeptin, 5 μ M pepstatin A, 1 mM PMSF, 1 mM DTT, and 3 μ l RiboLock™ (Thermo Fisher Scientific). The tubes were then sealed, mixed on a vortex mixer, and put on ice to cool down. The obtained homogenate was centrifuged at +4 °C for 7 minutes at maximum speed. The supernatant was carefully transferred to clean precooled tubes without disturbing the debris. The obtained lysate had a rather high detergent content, which could potentially cause dissociation of RNA-protein complexes and weaken the protein's bond with the substrate. To avoid this, 700 μ l of the buffer composed of 50 mM Tris-HCl (pH 7.4), 137 mM NaCl, 2.7 mM KCl was added to the lysate. At this stage, two control samples were collected: (1) lysate samples of 100 μ l were placed in separate tubes for future analysis of RNA input fraction; (2) lysate samples of 10 μ l were placed in 0.6 μ l tubes to evaluate the binding efficacy between the target protein and the substrate. Both types of tubes were stored at +4 °C.

Pull-down and wash of RNA-protein complexes

The tubes with the magnetic beads prepared earlier were placed on a magnetic stand, and the excess liquid was removed. The tubes were then removed from the stand, and the obtained diluted lysate was added to the beads. The tubes were placed on an orbital shaker, incubated at constant rotation for two hours at +4 °C, and transferred back to the magnetic stand. Lysate samples of 10 μ l were collected to control the protein's bond with the substrate, the remaining liquid was carefully removed.

The magnetic beads were washed by adding 400 μ l of the buffer composed of 50 mM Tris-HCl (pH 7.4), 137 mM NaCl, 2.7 mM KCl, and 0.05 % Igepal Ca-630 (Promega

Corp.). The tubes were gently shaken manually three times and placed on the magnetic stand. It should be noted that the number of washes is chosen for each RNA-protein complex on an individual basis. The assay described here included five washes. The tubes were incubated on an orbital shaker for five minutes at +4 °C during the last wash.

When the last wash was finished, the tubes were put on ice.

Evaluation of binding efficiency between protein and magnetic beads

At this stage, content, preservation, and binding efficiency of the target protein were evaluated. For this purpose, tubes with the previously collected lysate fractions described in sections "Tissue homogenization and collection of control samples" and "Pull-down and wash of RNA-protein complexes" were further analyzed. 1 μ l of 50 μ M HaloTag® TMR Ligand (Promega Corp.) was added to each tube. The content was mixed by pipetting, and the tubes were kept in the dark for 15 minutes. Then, 10 μ l of 4 \times SDS loading buffer was added and the mixture was heated for 2 minutes at +90 °C. We prepared 8 % polyacrylamide gel for Laemmli electrophoresis (Laemmli, 1970) and placed 5 μ l of the obtained product in gel wells. The gel was analyzed using a densitometric scanner Typhoon FLA 9000 (GE Healthcare) at the given wavelength (extinction wavelength of 532 nm and emission wavelength of 580 nm).

It can be seen from Figure 2 that a higher percentage of protein turned out to be bound and therefore was not detectable in the supernatant after 2-hour incubation. Otherwise, the go-to solution is to increase incubation time. A larger amount of magnetic beads increases non-specific binding.

RNA isolation

At this stage, two methods of isolating RNA from the protein complex are available, and prior knowledge of the studied protein, as well as the further course of analysis of the obtained RNA are to be taken into account.

For instance, isolating RNA directly from magnetic beads by incubation in the ExtractRNA reagent (Evrogen) (**the first method**) produces the eluate, which, in addition to the target RNA obtained directly from the protein, includes several non-

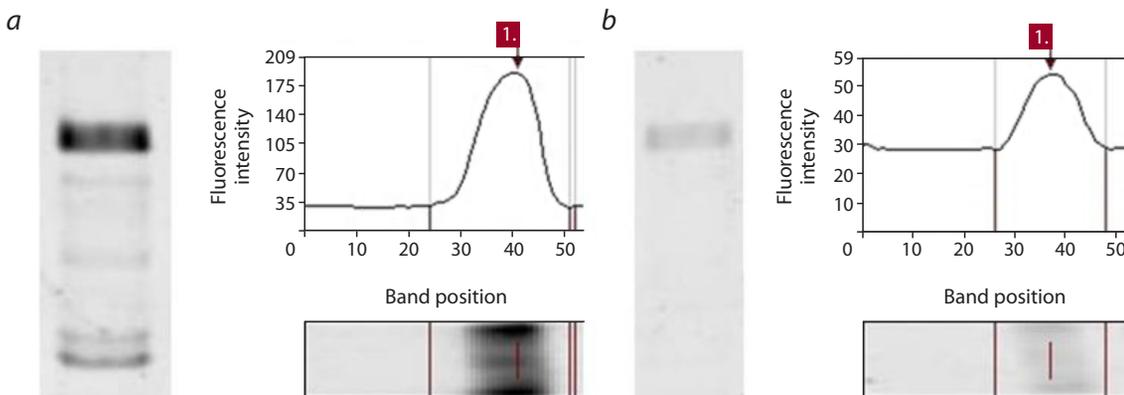


Fig. 2. Evaluation of the binding efficiency of the *EsCSDP3* protein on magnetic beads using the TMR ligand fluorescent dye.

a, The lysate sample before binding and the fluorescence level of the bound dye; b, the lysate sample 2 hours after incubation at +4 °C and the fluorescence level of the bound dye. The unbound protein fraction is about 25 %.

specifically bound RNA molecules from HaloTag® and the substrate. This isolation method is preferable, if the further analysis includes RT-PCR or Real-time PCR with primers on the known RNA targets.

If NGS is used to identify the nature of unknown RNA targets, then elution in TEV buffer and further extraction using the ExtractRNA reagent (Evrogen) (**the second method**) is recommended. TEV protease treatment of the RNA-protein complex facilitates its release into the solution, while non-specifically bound RNA molecules stay at the bottom of the tube.

To isolate RNA from the eluate, 100 µl of the buffer composed of 50 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 0.005 mM DTT, 40 U of RiboLock™ (Thermo Fisher Scientific) and 5 U of HaloTEV protease (Promega Corp.) was added to the washed beads with the precipitated target RNA-protein complex. The tube was placed on an orbital mixer and incubated overnight at +4 °C. The next day, the tubes were placed on a magnetic stand, and 90 µl of the eluate was transferred to a clean 1.5 ml tube. Then, 1 ml of the ExtractRNA reagent (Evrogen) was added to the obtained eluate. At this stage, RNA isolation from the beads and from the eluate proceeded identically. We similarly added 1 ml of the ExtractRNA reagent (Evrogen) to the tubes with magnetic beads washed in the buffer solution, and incubated the tubes on a magnetic stand at room temperature for 5 minutes with careful intermittent mixing. Then 200 µl of chloroform was added and the content was mixed on a vortex mixer for 30 seconds. The tubes were then centrifuged at +4 °C at 10,000g for 10 minutes. We carefully collected 500 µl of the aqueous phase and transferred it to a new tube. We then added 25 µg of glycogen, mixed the content by pipetting, and incubated the tube for 10 minutes at room temperature. The tubes were centrifuged for 10 minutes at 18,000g at room temperature. The supernatant was carefully removed with a small amount of isopropanol left at the bottom, and 1 ml of 75 % ethanol was added to the precipitate. The tubes were incubated at -20 °C overnight. Then the tubes were centrifuged at maximum speed at room temperature for

5 minutes. The supernatant was carefully removed, and the precipitate was dried for 10 minutes at room temperature and eluted into 20 µl of RNase-free water.

Analysis of the obtained RNA

To measure the concentration of the obtained RNA, a Quantus Fluorometer (Promega Corp.) was used. RNA profile was analyzed using a 2100 Bioanalyzer with RNA 6000 Nano and Pico kits (Agilent). Due to the high sensitivity of the device, a sample volume of 1 µl was sufficient for analysis, and therefore a sufficient amount of eluate may be preserved for further experiments.

The RNA profile obtained using the 2100 Bioanalyzer is presented in Figure 3. RNA obtained by elution from TEV buffer and further isolation using the ExtractRNA reagent (the second method) is shown in Figure 3, *a*. Comparison of the two RNA profiles shows that a much wider variety of various RNAs with higher RNA concentration was obtained by elution in the case of RNA-protein complex (plot 1), than in the case of HaloTag (plot 2). After more detailed consideration, we were able to notice that both samples had some common RNAs, and this should be taken into account in further analysis and comparisons.

The small RNA zone in samples from Figure 3, *a* (dashed outline) is shown in Figure 4. This zone was of special interest, since major differences between the samples were found. Its analysis also made it possible to predict the nature of RNA targets. The small RNA zone may be divided as follows: miRNA at nucleotide counts of ~40 and below; transfer RNAs at nucleotide counts of ~40 to ~80; small nuclear/nucleolar and ribosomal RNAs at nucleotide counts of ~80 to ~150.

As expected, comparison of the RNA profiles obtained using two different extraction methods showed a much higher fluorescence intensity in samples obtained using the first method (magnetic beads). In these samples, 5/5.8S ribosomal RNA made up the highest proportion of all RNA types. Comparison of the total fluorescence of the HaloTag and *EsCSDP3*

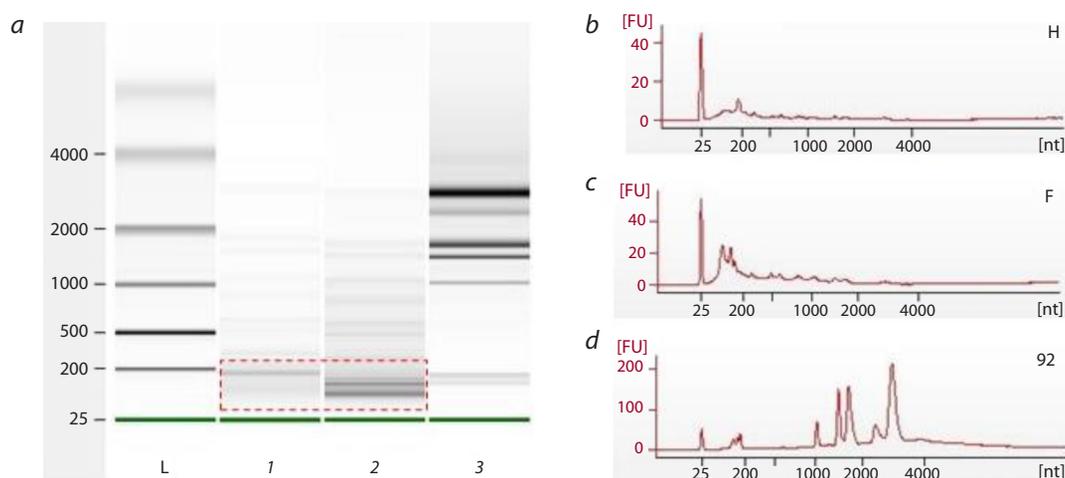


Fig. 3. RNA profile obtained using an Agilent 2100 Bioanalyzer with RNA 6000 Nano kit: *a–c*, RNA is isolated using the first method, where *a1* and *b* correspond to the HaloTag protein sample; *a2* and *c*, to the sample of the *EsCSDP3* protein; *a3* and *d*, to the total RNA sample used for reference.

A dashed outline indicates a small RNA zone. The RNA size compared to the reference marker is measured along the x-axis. The fluorescence intensity of the intercalating dye is measured along the y-axis.

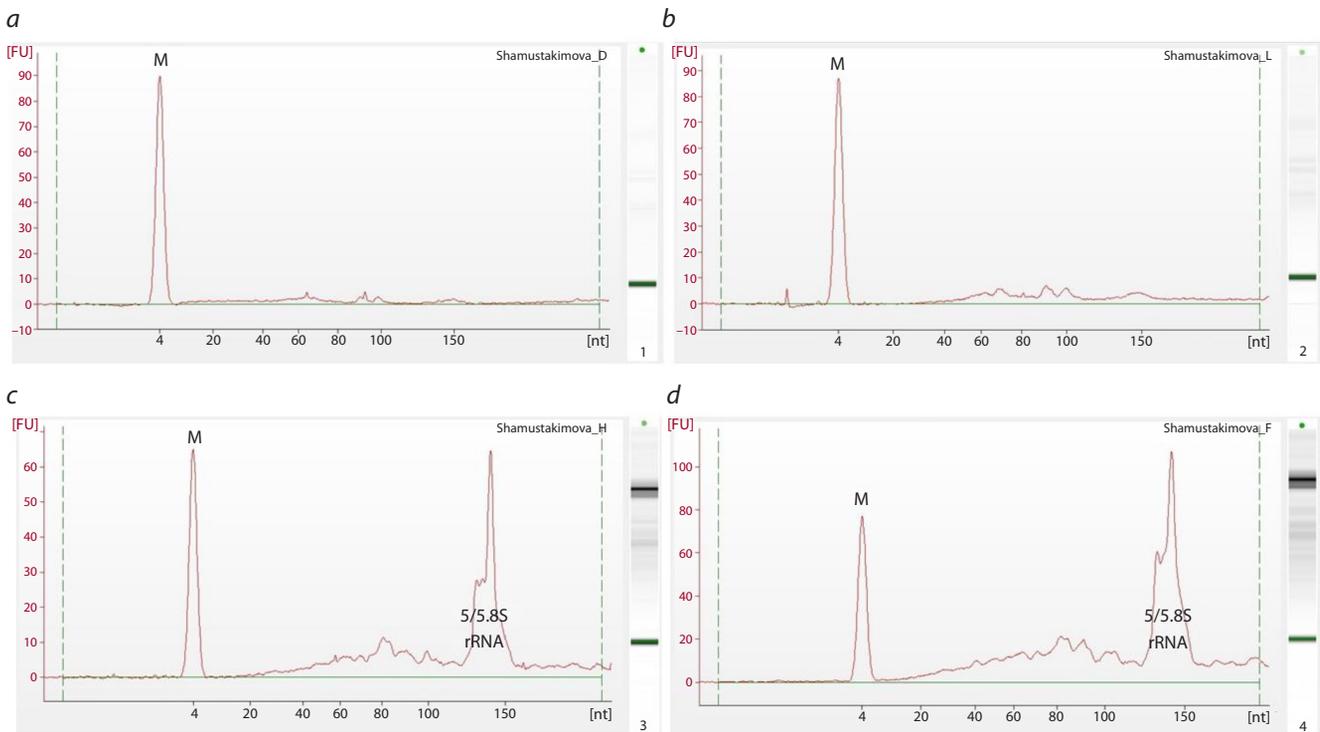


Fig. 4. RNA profile obtained using an Agilent 2100 Bioanalyzer with Small RNA kit for small RNA analysis: *a* and *b* are the HaloTag and *EsCSDP3* protein samples from the assay showed in Figure 3, *a* (dashed outline); *c* and *d* show the RNA isolated from the HaloTag and *EsCSDP3* protein samples using the second method.

The fluorescence intensity of the intercalating dye is measured along the y-axis. The size of the RNA is measured along the x-axis. M – lower alignment marker.

samples showed that the signal of the experimental sample (110 FU) was almost twice as intense as that of the control sample (65 FU).

Preparation of cDNA libraries and sequencing

The proposed protocol does not include a detailed description of cDNA library preparation and further sequencing procedures. It only lists the aspects to be taken into account in assay designs.

A cDNA library preparation kit should be selected based on the amount of RNA obtained. Although the *A. thaliana* genome is relatively short, its number of genes is comparable to that of humans, specifically 27,000 against 25,000. Thus, the recommended read depth is 15–30 million unpaired reads with lengths of 50 bp (Xing et al., 2015; Petri, Jakobsson, 2018). This read length is sufficient for mapping onto the genome, and the given read depth should be sufficient to identify specific targets based on their abundance quantitation in a statistically valid manner.

When RNA is obtained directly from magnetic beads, it is worth using a ribosomal RNA removal kit, since ribosomal RNA would account for a large number of reads due to their high abundance.

Conclusion

The obtained results have shown that the proposed method makes it possible to isolate complexes of fusion proteins and RNA targets from the *A. thaliana* leaves.

The Halo-RPD method has a number of advantages compared to the protocols based on immunoprecipitation, in par-

ticular, stable protein expression makes it possible to minimize both the initial amount of plant material for the assay and reagent consumption. The use of the reagent at the stage of RNA isolation from the eluate/substrate allows one to obtain and analyze small RNA and miRNA. The absence of covalent crosslinks and removal of the ultrasonic fragmentation stage make the proposed protocol applicable for analyzing native RNA profiles and thereby drawing preliminary conclusions on the nature of RNA targets, while the use of fluorescent dyes covalently bound to the HaloTag protein makes it possible to control the proper implementation of homogenization and protein isolation stages.

References

- Banks C.A.S., Boanca G., Lee Z.T., Eubanks C.G., Hattem G.L., Peak A., Weems L.E., Konkright J.J., Florens L., Washburn M.P. TNIP2 is a hub protein in the NF- κ B network with both protein and RNA mediated interactions. *Mol. Cell. Proteomics*. 2016;15(11):3435-3449. DOI 10.1074/mcp.M116.060509
- Brooks S.A., Rigby W.F.C. Characterization of the mRNA ligands bound by the RNA binding protein hnRNP A2 utilizing a novel *in vivo* technique. *Nucleic Acids Res*. 2000;28(10):e49. DOI 10.1093/nar/28.10.e49
- Frydrych Capelari É., da Fonseca G.C., Guzman F., Margis R. Circular and micro RNAs from *Arabidopsis thaliana* flowers are simultaneously isolated from AGO-IP libraries. *Plants*. 2019;8(9):302. DOI 10.3390/plants8090302
- Gu J., Wang M., Yang Y., Qiu D., Zhang Y., Ma J., Zhou Y., Hannon G.J., Yu Y. GoldCLIP: gel-omitted ligation-dependent CLIP. *Genom. Proteom. Bioinform.* 2018;16(2):136-143. DOI 10.1016/j.gpb.2018.04.003

- Köster T., Meyer K. Plant ribonemics: proteins in search of RNA partners. *Trends Plant Sci.* 2018;23(4):352-365. DOI 10.1016/j.tplants.2018.01.004
- Laemmli U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 1970;227(5259):680-685. DOI 10.1038/227680a0
- Li X., Pritykin Y., Concepcion C.P., Lu Y., La Rocca G., Zhang M., King B., Cook P.J., Au Y.W., Popow O., Paulo J.A., Otis H.J., Mastroleo C., Ogradowski P., Schreiner R., Haigis K.M., Betel D., Leslie C.S., Ventura A. High-resolution *in vivo* identification of miRNA targets by Halo-enhanced Ago2 pull-down. *Mol. Cell.* 2020; 79(1):167-179. DOI 10.1016/j.molcel.2020.05.009
- Los G.V., Encell L.P., McDougall M.G., Hartzell D.D., Karassina N., Zimprich C., Wood M.G., Learish R., Ohana R.F., Urh M., Simpson D., Mendez J., Zimmerman K., Otto P., Vidugiris G., Zhu J., Darzins A., Klaubert D.H., Bulleit R.F., Wood K.V. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* 2008;3(6):373-382. DOI 10.1021/cb800025k
- Petri R., Jakobsson J. Identifying miRNA targets using AGO-RIPseq. In: Lamandé S. (Ed.) mRNA Decay. Methods in Molecular Biology. Vol. 1720. New York: Humana Press, 2018;131-140. DOI 10.1007/978-1-4939-7540-2_9
- Ramanathan M., Porter D.F., Khavari P.A. Methods to study RNA-protein interactions. *Nat. Methods.* 2019;16(3):225-234. DOI 10.1038/s41592-019-0330-1
- Ren Z., Zhang D., Cao L., Zhang W., Zheng H., Liu Z., Han S., Dong Y., Zhu F., Liu H., Su H., Chen Y., Wu L., Zhu Y., Ku L. Functions and regulatory framework of ZmNST3 in maize under lodging and drought stress. *Plant Cell Environ.* 2020;43(9):2272-2286. DOI 10.1111/pce.13829
- Samanta S., Thakur J.K. Characterization of mediator complex and its associated proteins from rice. In: Kaufmann K., Mueller-Roeber B. (Eds.) Plant Gene Regulatory Networks. Methods in Molecular Biology. Vol. 1629. New York: Humana Press, 2017;123-140. DOI 10.1007/978-1-4939-7125-1_9
- Seo J.S., Chua N.H. Analysis of interaction between long noncoding RNAs and protein by RNA immunoprecipitation in *Arabidopsis*. In: Chekanova J.A., Wang H.-L.V. (Eds.) Plant Long Non-Coding RNAs. Methods in Molecular Biology. Vol. 1933. New York: Humana Press, 2019;289-295. DOI 10.1007/978-1-4939-9045-0_18
- Sorenson R., Bailey-Serres J. Rapid immunopurification of ribonucleo-protein complexes of plants. In: Alonso J., Stepanova A. (Eds.) Plant Functional Genomics. Methods in Molecular Biology. Vol. 1284. New York: Humana Press, 2015;209-219. DOI 10.1007/978-1-4939-2444-8_10
- Steffen A., Elgner M., Staiger D. Regulation of flowering time by the RNA-binding proteins AtGRP7 and AtGRP8. *Plant Cell Physiol.* 2019;60(9):2040-2050. DOI 10.1093/pcp/pcz124
- Taranov V.V., Zlobin N.E., Evlakov K.I., Shamustakimova A.O., Babakov A.V. Contribution of *Eutrema salsugineum* cold shock domain structure to the interaction with RNA. *Biochemistry (Moscow).* 2018;83(11):1369-1379. DOI 10.1134/S000629791811007X
- Urh M., Hartzell D., Mendez J., Klaubert D.H., Wood K. Methods for detection of protein-protein and protein-DNA interactions using HaloTagTM. In: Zachariou M. (Ed.) Affinity Chromatography. Methods in Molecular Biology. Vol. 421. New York: Humana Press, 2008;191-210. DOI 10.1007/978-1-59745-582-4_13
- van Dijk M., Visser A., Buabeng K.M., Poutsma A., van der Schors R.C., Oudejans C.B. Mutations within the *LINC-HELLP* non-coding RNA differentially bind ribosomal and RNA splicing complexes and negatively affect trophoblast differentiation. *Hum. Mol. Genet.* 2015; 24(19):5475-5485. DOI 10.1093/hmg/ddv274
- Xing D., Wang Y., Hamilton M., Ben-Hur A., Reddy A.S. Transcriptome-wide identification of RNA targets of Arabidopsis SERINE/ARGININE-RICH45 uncovers the unexpected roles of this RNA binding protein in RNA processing. *Plant Cell.* 2015;27(12):3294-3308. DOI 10.1105/tpc.15.00641

ORCID

A.O. Shamustakimova orcid.org/0000-0003-3535-3108

Acknowledgements. The author thanks Charles Banks from the Stowers Institute for Medical Research for assistance in protocol development. The work was supported by the Russian Foundation for Basic Research (project Nos. 05-04-89005-NWO and 14-04-00816). The equipment was provided by the Common Use Center "Biotechnology" of the All-Russian Research Institute of Agricultural Biotechnology (project No. RFMEFI62114X0003).

Conflict of interest. The author declares no conflict of interest.

Received March 28, 2023. Revised August 21, 2023. Accepted August 24, 2023.

Перевод на английский язык <https://vavilov.elpub.ru/jour>

Associations of CAG repeat polymorphism in the androgen receptor gene with steroid hormone levels and anthropometrics among men: the role of the ethnic factor

L.V. Osadchuk , G.V. Vasiliev, A.V. Osadchuk

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
 losadch@bionet.nsc.ru

Abstract. Androgens are required for stimulation and maintenance of skeletal growth and bone homeostasis. Physiological functions of androgens are mediated through the androgen receptor (AR). The androgen receptor gene AR has a polymorphic trinucleotide CAG repeat and the length of AR CAG repeats determining the sensitivity of bone tissue to androgens is associated with skeleton formation and body proportions. This study aimed to investigate the relationship between AR CAG repeat polymorphism, circulating sex steroid hormones and the anthropometrics in males of different ethnic origins. Male volunteers of three ethnic groups (Slavs, Buryats, Yakuts) from urban Russian populations were recruited in a population-based study ($n = 1078$). Anthropometric indicators (height, arm span, leg length, the length of 2 and 4 digits of both hands) were measured and the following anthropometric indices were calculated: the ratio of height to leg length, the ratio of arm span to height, the ratio of lengths of second to fourth digit of the hand. Serum testosterone and estradiol were determined by enzyme immunoassay. Genotyping of the AR CAG repeats was performed using fragment analysis and capillary electrophoresis. Ethnic differences in all anthropometric and hormonal indicators have been established, with higher anthropometric indicators in Slavs than Buryats, and in most cases higher than in Yakuts. The testosterone level was higher among Slavs compared to Buryats, but did not differ from Yakuts; the estradiol level was lower among Slavs compared to Buryats, who did not differ from Yakuts. Buryats and Yakuts had a higher number of CAG repeats than Slavs (medians: Slavs, 23; Buryats, 24; Yakuts, 25). Positive correlations were found between the length of AR CAG repeats and estradiol levels in Buryats and testosterone levels in Yakuts, while longer CAG repeats were accompanied by higher estradiol levels in Buryats and testosterone levels in Slavs and Yakuts. Ethnic-specific correlations have been established between the steroid hormone levels and some anthropometric indicators in all ethnic groups. Available data suggest that the ethnic-specific associations of AR CAG repeats with anthropometrics can be mediated by sex steroid hormones as important regulators of skeletal growth and bone homeostasis.
Key words: AR CAG repeat polymorphism; testosterone; estradiol; anthropometrics; human male population.

For citation: Osadchuk L.V., Vasiliev G.V., Osadchuk A.V. Associations of CAG repeat polymorphism in the androgen receptor gene with steroid hormone levels and anthropometrics among men: the role of the ethnic factor. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(1):80-89. DOI 10.18699/vjgb-24-10

Ассоциации САG-полиморфизма гена андрогенового рецептора с уровнем стероидных гормонов и антропометрическими показателями у мужчин: роль этнического фактора

Л.В. Осадчук , Г.В. Васильев, А.В. Осадчук

Федеральный исследовательский центр «Институт цитологии и генетики Сибирского отделения Российской академии наук», Новосибирск, Россия
 losadch@bionet.nsc.ru

Аннотация. Андрогены необходимы для роста и поддержания гомеостаза костной ткани. Физиологические функции андрогенов опосредованы андрогеновым рецептором (AR). Ген андрогенового рецептора AR имеет полиморфный тринуклеотидный CAG-повтор, и длина AR CAG-повторов, регулируя чувствительность костной ткани к андрогенам, оказывает влияние на формирование скелета и пропорций тела. Цель исследования – выявить ассоциации между длиной AR CAG-повторов, гормональными и антропометрическими показателями у мужчин различных этнических групп. В популяционном исследовании приняли участие мужчины добровольцы ($n = 1078$) трех этнических групп (славян, бурят и якутов) из городских популяций России. У участников измеряли рост, размах рук, длину ноги, длину второго и четвертого пальцев обеих рук, рассчитывали ростовые индексы: отношение роста к длине ноги, отношение размаха рук к росту, отношение длины второго к четвертому пальцу. Тестостерон и эстрадиол в образцах периферической крови определяли иммуноферментным методом. Генотипирование AR CAG-повторов проводили с помощью фрагментного анализа и капиллярного электрофореза.

Установлены более высокие антропометрические показатели у славян по сравнению с бурятами и в большинстве случаев с якутами. Уровень тестостерона повышен у славян по сравнению с бурятами, но не отличался от якутов, а уровень эстрадиола понижен по сравнению с таковым у бурят или якутов, которые не отличались по этому показателю. Длина AR CAG-повторов составила 23, 24 и 25 триплетов (медианы) у славян, бурят и якутов соответственно. Выявлены положительные корреляции между длиной AR CAG-повторов и уровнем эстрадиола у бурят и уровнем тестостерона у якутов, причем длинные CAG-повторы сопровождалась повышенным уровнем эстрадиола у бурят и тестостерона у славян и якутов. Установлены этнозависимые корреляции между уровнем стероидных гормонов и антропометрическими показателями у всех этносов. Полученные данные предполагают существование этнозависимых ассоциаций AR CAG-полиморфизма с размерами костей скелета, которые опосредуются стероидными гормонами как важными регуляторами роста и гомеостаза костной ткани.

Ключевые слова: полиморфизм AR CAG-повторов; антропометрия; тестостерон; эстрадиол; этнические различия; популяции человека.

Introduction

Androgens, which are secreted by the Leydig cells of the testes, play a critical role in normal male physiology, and impairment of the androgen action on the target tissue is accompanied by a wide range of pathological changes. The main role of androgens consists in ontogenetic formation and maintenance of the male phenotype integrity, including a normal process of sexual differentiation, pubertal development, formation and maintenance of secondary sex characteristics, sexual behavior, and sperm production. Besides reproductive effects, androgens affect the functions of non-reproductive tissues, in particular, development and growth of the skeletal system, formation of stature and body proportions (Zitzmann, Nieschlag, 2003; Almeida et al., 2017; Alemany, 2022). Androgens exert their physiological effects on bone growth and maintenance of bone metabolism together with estrogens, which bind to the estrogen receptors present in bone tissue (Almeida et al., 2017; Alemany, 2022).

The process of bone tissue formation has several age stages. The first of them takes place *in utero* and is under the control of sex steroid hormones. Since the estimation of fetal androgens is complex, most investigators suggest using biomarkers, which are stable indicators reflecting the degree of exposure to androgens during fetal development. The ratio between length of the index and ring digit (2D:4D) has therefore been proposed to serve as a proxy marker for *in utero* androgen exposure. Several reviews and meta-analyses have shown that in most cases men and boys have lower values of this ratio than women and girls, suggesting that the gender difference in the 2D:4D ratio is determined by higher levels of androgens in male embryos compared to female (Hönekopp et al., 2007; Grimbois et al., 2010; Knickmeyer et al., 2011; Xu, Zheng, 2015; Swift-Gallant et al., 2020; Jägetoft et al., 2022). One study (Mitsui et al., 2015) showed gender differences in testosterone levels of umbilical cord blood samples, which were significantly higher in samples collected from males than those from females. These data confirm the hypothesis of sexual dimorphism with respect to the 2D:4D ratio, which reflects the levels of embryonic sex steroids. Sexual dimorphism in relation to the 2D:4D ratio persists throughout life, although data concerning sexual differences in this ratio in childhood are less variable compared to data in adults (Knickmeyer et al., 2011; Mitsui et al., 2015; Jägetoft et al., 2022).

The link found between the finger development and the prenatal androgens (and estrogens) suggests that the 2D:4D

ratio is inversely related to prenatal testosterone levels and positively related to prenatal estrogen levels. There is some evidence that the 2D:4D ratio shows associations with sex steroid hormones in adults, so it can be used as a proxy marker of embryonic effects of sex steroids on a number of physiological, behavioral and anthropometric traits expressed in adults (Knickmeyer et al., 2011; Manning et al., 2014). However, the 2D:4D ratio is not always associated with the sex steroid levels in the adult male population, so the digit ratio is often the subject of debate about causality and validity as an indicator of the embryonic androgen level (Richards et al., 2020; Swift-Gallant et al., 2020).

Puberty is a unique stage of postnatal development characterized by substantial anatomical and physiological changes leading to the accumulation of bone mass, bone growth in length and the formation of a stature that is controlled by reactivation of the hypothalamic-pituitary-gonadal axis after a long period of quiet. At the beginning of puberty, androgens together with estrogens stimulate the pubertal growth spurt by increasing growth hormone (GH) secretion and hepatic insulin-like growth factor-1 (IGF-I) release, but sex steroids also have a direct effect on bone growth, since there are androgen and estrogen receptors in chondrocytes. At the end of puberty, high estrogen concentrations, but not androgens, block the longitudinal growth of bones in boys, stimulating the closure of epiphyseal growth plates, an effect mediated by the direct action of estrogens on proliferating chondrocytes (Almeida et al., 2017). Androgens are known to be precursors of estrogens in the synthetic pathways of the steroid hormones. In men, about 15 % of estradiol is secreted directly from the testes, and the remaining 85 % is derived from peripheral aromatization of androgens to estrogens by the aromatase enzyme (CYP19A1) (Almeida et al., 2017; Alemany, 2022). Estrogen resistance or aromatase deficiency in male adolescents leads to a delay in bone age and high growth, despite normal or high testosterone concentrations (Frank, 2003; Alemany, 2022).

Androgens, as well as all steroid hormones, do not affect target tissues immediately, but perform their effects essentially in the medium term, modulating gene expression; their action is relatively long and is regulated by a complex network of adaptive mechanisms in accordance with the needs of the body. Most physiological effects of androgens are mediated through the androgen receptor (AR). Since ARs are expressed in almost every tissue, "androgenicity" is manifested almost everywhere; therefore, the role of androgen receptors in males

is fundamentally important (Zitzmann, Nieschlag, 2003). The androgen receptor belongs to the family of nuclear receptors of steroid and thyroid hormones and, like other members of the family, is able to interact directly with nuclear DNA. AR is a ligand-dependent nuclear transcription factor, which is activated when binding to androgens (testosterone and dihydrotestosterone) and changes the expression of AR-dependent target genes (Davey, Grossmann, 2016; Xiao et al., 2016). Interacting with certain regulatory regions, AR serves as a transcription factor regulating the synthesis of a number of proteins. There are also non-genomic effects of AR unrelated to gene expression regulation.

The AR gene has a trinucleotide polymorphic CAG repeat (cytosine–adenine–guanine) in exon 1 that is transcribed into a different number of polyglutamine amino acids; that is, the variability in AR size is partially due to this trinucleotide repeat. In healthy males, the normal range of CAG repeats is 11–31 triplets, and transactivation activity of AR is inversely proportional to the number of CAG repeats (Davey, Grossmann, 2016). *In vitro* and *in vivo* studies have shown that the longer the length of CAG repeats, the weaker the transactivation ability of AR and the weaker the effects of androgens in target tissues (Buchanan et al., 2004). The authors suggest that normal function of AR is maintained in a critical and limited range of CAG repeats (16–29 triplets); the number of CAG repeats outside this range can be associated with impaired function of androgen-dependent tissues and various diseases (Davis-Dao et al., 2007; Davey, Grossmann, 2016; Ryan et al., 2017; Wang et al., 2018; Osadchuk L., Osadchuk A., 2022). It should also be noted that the testosterone effects appear after it binds to AR, affecting the transcriptional activity of AR. Thus, the AR CAG polymorphism, through reducing the sensitivity of target tissues to androgens, weakens the physiological effects of androgens and therefore is a crucial factor determining the “masculinity” of every man.

Androgen receptors are expressed in all types of bone cells (osteoblasts, osteoclasts and osteocytes). Since some polymorphic variants of AR modulate sensitivity to androgens, differences in the association between the testosterone levels and bone mass may be associated with CAG polymorphism in the AR gene. It has been established that an increased number of CAG repeats in the AR gene attenuates the testosterone effects on bone density and bone metabolism (Zitzmann, Nieschlag, 2003). In addition, a relationship was revealed between the AR CAG polymorphism and bone mineral content and bone mineral density, which was modulated by the free testosterone level (Guadalupe-Grau et al., 2010). The longitudinal growth was inversely associated with the AR CAG repeat length in boys from early prepubertal to pubertal age, but this association diminished in subsequent years and completely disappeared after the age of 16 years. The height of adult males was not associated with the AR CAG repeat length (Voorhoeve et al., 2011). The authors believe that during puberty, this relationship disappears, possibly due to the compensative increase in the activity of the hypothalamic–pituitary–gonadal axis. The effects of the AR CAG repeat length on bone mass were investigated in prepubertal boys of 12 years old (Rodríguez-García et al., 2015). In boys with longer AR CAG alleles, height, body mass, bone mineral

density and content were increased compared to boys with shorter AR CAG alleles, which confirms the hypothesis that longer AR CAG alleles are associated with an increase in bone mass in prepubertal boys.

As already mentioned, we can expect a close relationship between the circulating testosterone level and the AR CAG polymorphism. Most studies have shown that the CAG repeat length directly correlates with serum testosterone levels in adult men (Crabbe et al., 2007; Huhtaniemi et al., 2009; Ma et al., 2014; Grigorova et al., 2017; Khan et al., 2018). The authors believe that longer CAG repeats impair androgen feedback in the hypothalamic–pituitary–testicular system and, thus, can increase testosterone levels. The weaker transcriptional activity of AR with a longer CAG repeat length seems to be totally or nearly totally compensated for by higher testosterone levels; therefore, an increase in testosterone levels can be considered as a compensatory effect to maintain an adequate androgen status of a man. From a genetic perspective, the testosterone level is undoubtedly a polygenic trait, and the AR CAG polymorphism is just one of many genetic factors underlying the genetic control of this trait. Individual variation in testosterone levels in healthy men can be explained by the AR CAG polymorphism. In a large study population of healthy Belgian men, it was shown that the CAG repeat length was positively associated with serum total testosterone (Crabbe et al., 2007). The authors suggest that in men, 6.0–8.5 % of individual testosterone variability can be associated with the CAG repeat length in the AR gene.

In our previous work, it was shown that the variability of the AR CAG repeat length was associated with the ethnic composition of the studied population (Osadchuk et al., 2022). Significant differences were observed in the AR CAG repeat length between the most common ethnic cohorts of Slavs (Caucasians), Buryats (Asians), and Yakuts (Asians) with median in Slavs – 23; Buryats – 24; Yakuts – 25. The Slavs have the largest range (7–36 repeats), the Yakuts have the smallest range (18–32 repeats) and the Buryats have the middle range (11–39 repeats). Longer CAG repeats were associated with an impaired semen quality within the Slav and Buryat groups, but this effect was not found in Yakuts.

Based on the above, we can expect there to be an effect of the AR CAG polymorphism on the variability of the sex steroid hormone levels in Russian men, which may affect the formation of some androgen-dependent anthropometric parameters. The aim of this study was to study the possible relationship between variations in the AR CAG repeat length, which affects the transactivation activity of AR, and circulating sex steroid hormones, as well as with steroid-dependent anthropometric parameters. In addition, it was interesting to compare the above associations in men of different ethnicity, who also differ in the AR CAG repeat length, anthropometric indicators, testosterone and estradiol levels. To achieve these goals, a multicenter population study of Russian male volunteers was conducted, in which the AR CAG repeat length, serum testosterone and estradiol levels, anthropometric indicators, including height, leg length, arm span, the 2nd to 4th digit ratio of both hands were determined. To clarify the role of the ethnic factor, the study was conducted on men of three ethnic groups – Slavs, Buryats and Yakuts, who had previously been

studied with respect to the AR CAG polymorphism (Osadchuk et al., 2022). It should be noted that this is the first Russian study clarifying the role of the AR CAG polymorphism in regulating sex steroid hormone levels and steroid-dependent anthropometric indicators.

Materials and methods

Young male volunteers ($n = 1078$) of three ethnicities from five Russian cities participated in this study: Slavs from Archangelsk, Novosibirsk, Kemerovo; Buryats from Ulan-Ude, Yakuts from Yakutsk. The Slavic group consisted of Russians (95.9 %), Ukrainians (0.5 %) and descendants of mixed marriages of Russians, Belarusians, Ukrainians (3.5 %). The men filled out questionnaires, including questions about age, place of birth, nationality, profession, work conditions, noted military service, smoking and alcohol consumption, past and current diseases. Ethnicity of participants was obtained according to the information from the self-reporting questionnaires, taking into account ethnicities of the participants' parents and grandparents. As a rule, the participants considered themselves healthy and had not previously consulted doctors about chronic diseases. The inclusion criteria for participation in the study were the absence of acute diseases or taking medications that affect hormone levels or bone metabolism. All participants gave informed consent for participation. The ethics committee of the Federal Research Center "Institute of Cytology and Genetics", the Siberian Branch of the Russian Academy of Sciences, approved the study (Protocol No. 160, date 17.09.2020).

Height, body weight, waist and hip circumference (WC and HC, respectively), leg length, arm span, length of index and ring fingers of both hands were measured in all participants. Body mass index (BMI), trochanteric index (TI), androgenic deficiency index (ADI) and digital index (2D:4D right and left) were calculated. Body weight was estimated in kg, WC, HC, height, leg length, arm span in cm. BMI is the main anthropometric indicator of obesity and is calculated as the ratio of body weight (kg)/height (m^2). TI characterizes the body proportion formed by the end of puberty and is calculated as the ratio of height to leg length. ADI also characterizes the proportions of the body formed by the end of puberty and is calculated as the ratio of arm span to height. This indicator depends on the testosterone level in adolescents, and if androgen deficiency took place before puberty, then the arm span begins to exceed the height and a specific tallness with eunuchoid body proportions is formed (Frank, 2003). The length of 2D and 4D was measured by digital caliper; the 2D:4D ratio was calculated as the ratio of the length of 2 finger to 4 finger. It is assumed that the 2D:4D ratio reflects the prenatal androgenization and does not change during postnatal life (Knickmeyer et al., 2011; Manning et al., 2011).

Fasting blood samples from the cubital vein were drawn in the morning between 8–11 hr. Blood samples were centrifuged in 15–20 min and at 1500 rpm; serum was stored at $-40\text{ }^\circ\text{C}$ until hormonal analysis. Serum concentrations of luteinizing hormone (LH), follicle-stimulating hormone (FSH), testosterone (T); estradiol (E_2) were determined by enzyme immunoassay with commercially available kits (Alkor Bio, Xema, Russia) according to the manufacturer manuals. The ranges

of evaluated concentrations for our study population were as follows: LH – 1.3–6.7 mIU/ml; FSH – 1.3–8.8 mIU/ml; T – 11.7–38.2 nmol/L; E_2 – 0.10–0.35 nmol/L.

This study included men who had previously been genotyped for AR CAG repeats; the genotyping technique was described in detail earlier (Osadchuk et al., 2022). Briefly, genomic DNA was extracted from peripheral blood leukocytes using the common phenol-chloroform method. Genotyping of the AR CAG repeats was performed by the method of fragment PCR analysis and capillary electrophoresis using the sequencer "Nanophor-05" (Syntol, Russia). The method allows to determine the relative length of the product in relation to the length standard and is based on the separation of DNA into fractions by molecular weight. The primer sequences were as follows: forward, 5'-(FAM)-TCCAGAATCTGTTCCAGAGCGTGC-3' and reverse, 5'-GCTGTGAAGGTTGCTGTTCCCTCAT-3'. The number of CAG repeats was calculated using an allelic ladder of marker fragments, which consisted of eight fragments of different lengths and were used as an internal standard for calculations (lengths of CAG repeats were 12, 19, 21, 23, 25, 27, 29, 33 triplets).

The statistical analysis of the data was performed using the statistical package "STATISTICA" (version 8.0). The results are presented as mean (SD). The Kolmogorov-Smirnov test for normality was used. Since most parameters were not normally distributed, the Kruskal-Wallis ANOVA test for comparing multiple independent groups was carried out to find the differences in the hormonal and anthropometric parameters among ethnic groups. Spearman correlation coefficients were used to determine correlations among parameters in each ethnic group. Duncan's test was used for pairwise comparison of groups. The testosterone and estradiol levels were best normalized by log transformation before analysis.

The next step to find out possible associations between the AR CAG polymorphism and hormonal and anthropometric indicators was the stratification of participants in each ethnic group into three CAG categories based on the CAG ethnic range restriction with a frequency below 5 % for short or long CAG repeat length. The rest of the CAG repeat range represented the medium CAG repeat category. The categorization of the CAG repeat length corresponded to the one presented earlier (Osadchuk et al., 2022). The results of the stratification of participants by the CAG repeat length (number of triplets) are as follows: Buryats – short (≤ 20); medium (21–27); long (≥ 28); Slavs – short (≤ 19); medium (20–24); long (≥ 25); Yakuts – short (≤ 22); medium (23–27); long (≥ 28). Differences in anthropometric and hormonal variables between subgroups with different CAG repeat length were tested by analysis of covariance (ANCOVA); hormonal variables were adjusted for age, body weight, WC, BMI. A p -value ≤ 0.05 was regarded as statistically significant.

Results

Ethnic differences in anthropometric and hormonal parameters, and the AR CAG repeat length. The ethnic groups differed in age; the Buryats were 1.5 years younger than the Slavs and the Yakuts ($p \leq 0.05$), who did not differ from each other (Table 1). Almost all anthropometric indicators were noted to be higher in Slavs compared to Buryats ($p \leq 0.05$),

Table 1. Anthropometric and hormonal parameters, and AR CAG repeat length in men of three ethnic groups

Parameter	Buryats, <i>n</i> = 223	Slavs, <i>n</i> = 708	Yakuts, <i>n</i> = 147
Age, years	23.7 (6.5) ^a	25.3 (6.6) ^b	25.4 (7.3) ^b
Weight, kg	70.9 (13.2) ^a	78.1 (13.9) ^b	70.6 (13.8) ^a
Height, cm	174.8 (6.3) ^a	179.0 (6.9) ^b	172.2 (6.5) ^c
Waist circumference, cm	82.0 (12.1) ^a	84.8 (10.6) ^b	84.6 (12.7) ^b
Hip circumference, cm	94.2 (7.3) ^a	99.1 (8.2) ^b	97.2 (7.9) ^c
BMI, kg/m ²	23.2 (4.3) ^a	24.3 (3.9) ^b	23.8 (4.4) ^{ab}
Leg length, cm	89.2 (5.1) ^a	94.2 (5.3) ^b	89.4 (5.5) ^a
TI	1.96 (0.07) ^a	1.90 (0.08) ^b	1.93 (0.08) ^c
Arm span, cm	177.8 (7.4) ^a	182.3 (8.1) ^b	177.7 (7.4) ^a
ADI	1.017 (0.025) ^a	1.018 (0.029) ^a	1.032 (0.021) ^b
Index digit right, cm	7.7 (1.1) ^a	7.8 (1.0) ^b	7.8 (0.9) ^{ab}
Index digit left, cm	7.7 (1.1) ^a	7.8 (1.0) ^b	7.7 (0.9) ^{ab}
Ring digit right, cm	8.1 (1.2) ^a	8.0 (1.1) ^b	8.2 (1.0) ^a
Ring digit left, cm	8.1 (1.2) ^a	8.0 (1.1) ^b	8.2 (1.0) ^a
2D:4D right	0.95 (0.03) ^a	0.98 (0.04) ^b	0.95 (0.03) ^a
2D:4D left	0.95 (0.03) ^a	0.98 (0.04) ^b	0.94 (0.04) ^a
LH, mIU/ml	3.96 (1.66) ^a	3.51 (1.53) ^b	3.58 (1.56) ^b
FSH, mIU/ml	4.76 (3.12) ^a	3.76 (2.76) ^b	5.31 (3.24) ^a
Testosterone, nmol/L	18.81 (5.98) ^a	21.16 (7.58) ^b	19.67 (6.59) ^{ab}
Estradiol, nmol/L	0.227 (0.059) ^a	0.194 (0.067) ^b	0.224 (0.055) ^{ab}
T/E ₂	87.3 (37.0) ^a	120.7 (62.9) ^b	91.7 (35.1) ^a
CAG repeat number	24.0 (3.5) ^a	23.0 (3.1) ^b	25.0 (2.8) ^c

Note here and further. Values are presented as mean (SD); BMI, body mass index; TI, trochanter index (the ratio of height to leg length); ADI, androgen deficiency index (the ratio of the length of arm span to height); 2D:4D-right, 2D:4D-left (the ratio of the length of the second finger to fourth); LH, luteinizing hormone; FSH, follicle-stimulating hormone; T/E₂, the ratio of testosterone to estradiol concentrations; a, b, c comparisons with different superscripts within variables were significant ($p \leq 0.05$).

with the exception of lower values of TI and the length of the ring finger in Slavs ($p \leq 0.05$). The Slavs differed from the Yakuts by higher values of almost all anthropometric indicators ($p \leq 0.05$), but did not differ in WC and length of both index fingers, and were characterized by lower TI, ADI and the length of both ring fingers ($p \leq 0.05$, see Table 1). The Buryats did not differ from the Yakuts in most anthropometric indicators, with the exception of higher height, TI and lower WC and ADI ($p \leq 0.05$). Thus, the Buryats and the Yakuts are closer to each other in body proportions and are significantly differentiated from the Slavs.

The serum levels of LH, FSH and estradiol were lower, and the testosterone level and the T/E₂ ratio were significantly higher in Slavs compared to Buryats ($p \leq 0.05$, see Table 1). Buryats and Yakuts did not differ from each other in all hormonal parameters, with the exception of lower LH values in Yakuts ($p \leq 0.05$). Thus, the Buryats and the Yakuts were close

to each other in hormonal status and more differentiated from the Slavs. The length of CAG repeats significantly differed between all ethnic groups (medians: Slavs – 23; Buryats – 24; Yakuts – 25, $p \leq 0.05$).

Correlations between the AR CAG repeat length, hormonal and anthropometric indicators in men of three ethnic groups. The relationships between the CAG repeat length, hormonal and anthropometric indicators were determined by ethnic origin (Table 2). In Buryats, a positive correlation was detected between the CAG repeat length and the 2D:4D ratio right, as well as the estradiol level ($p \leq 0.05$), in Yakuts – between the CAG repeat length and the testosterone level ($p \leq 0.05$), and in Slavs, no reliable correlations were found (see Table 2). More numerous correlations were observed between hormone levels and anthropometric characteristics, which were also modulated by ethnic origin (see Table 2). In Buryats, negative correlations were established

Table 2. Correlations between anthropometric and hormonal parameters, and AR CAG repeat length in men of three ethnic groups (Spearman's test)

Parameter	Buryats, n = 223			Slavs, n = 708			Yakuts, n = 147		
	T	E ₂	CAG repeats	T	E ₂	CAG repeats	T	E ₂	CAG repeats
Age, years	-0.139*	-0.300*	-0.081	-0.297*	-0.215*	-0.034	-0.123	0.014	-0.130
Weight, kg	-0.341*	-0.153*	-0.090	-0.384*	-0.047	-0.027	-0.409*	0.067	-0.040
Height, cm	0.110	0.021	-0.093	-0.019	-0.003	-0.015	0.069	0.156	0.010
Leg length, cm	0.121	-0.003	-0.009	0.094*	0.010	-0.027	-0.036	0.247*	0.054
TI	-0.098	0.066	-0.092	-0.151*	0.001	0.032	0.065	-0.251*	-0.068
Arm span, cm	0.083	-0.141*	-0.032	-0.004	0.035	-0.001	0.105	0.068	0.028
ADI	-0.035	-0.262*	0.063	0.011	0.046	0.034	0.079	-0.146	0.101
2D-right, cm	0.125	-0.015	0.044	-0.177*	0.213*	-0.007	0.083	0.195*	-0.015
2D-left, cm	0.122	-0.035	0.019	-0.179*	0.189*	-0.009	0.053	0.213*	0.014
4D-right, cm	0.089	-0.037	-0.033	-0.212*	0.198*	-0.026	0.086	0.202*	-0.016
4D-left, cm	0.087	-0.059	-0.019	-0.204*	0.208*	-0.009	0.064	0.201*	-0.014
2D:4D-right	0.109	0.108	0.189*	0.093*	0.006	0.044	-0.010	0.020	0.046
2D:4D-left	0.103	0.097	0.119	0.097*	-0.038	-0.013	-0.046	0.025	0.086
LH, mIU/ml	0.173*	0.085	-0.010	0.120*	-0.027	0.043	0.090	0.184*	0.052
Testosterone, nmol/L	-	0.178*	0.090	-	0.044	0.062	-	0.109	0.164*
Estradiol, nmol/L	0.178*	-	0.147*	0.044	-	-0.030	0.109	-	-0.037

* – Correlation is significant ($p < 0.05$).

between estradiol level and arm span, and ADI ($p \leq 0.05$); in Yakuts – positive correlations between estradiol level and leg length, and lengths of all four digits ($p \leq 0.05$), but negative correlation between the estradiol level and TI ($p \leq 0.05$). The Slavs had a positive correlation between testosterone level and leg length, as well as both 2D:4D ratios ($p \leq 0.05$); a negative correlation between testosterone level and TI, and lengths of all four digits ($p \leq 0.05$). In addition, positive correlations were noted between estradiol level and lengths of all four digits (see Table 2, $p \leq 0.05$).

Anthropometric and hormonal parameters in subgroups with different AR CAG repeat lengths in men of three ethnic groups. Anthropometric and hormonal data in the subgroups of short, medium and long CAG repeats in each ethnic group are presented in Table 3. The Buryats had significant differences in height, ADI, the 2D:4D ratio right, estradiol level between the subgroups with short and long CAG repeats ($p \leq 0.05$), the Slavs – in leg length and testosterone level ($p \leq 0.05$), the Yakuts – in arm span, ADI, testosterone level ($p \leq 0.05$). Thus, in all ethnic groups, long CAG repeats were accompanied by an increased level of steroid hormones.

Discussion

Comparison of anthropometric and hormonal indicators in young Russian men of different ethnicity allowed us to establish reliable ethnic differences in various anthropometric

indicators and the levels of sex steroid hormones. Attention is drawn to the higher anthropometric indicators that determine the male stature (height, leg length, arm span) in the Slavs than the Buryats, and in most cases, than the Yakuts, which corresponds to a higher current testosterone level and a lower current estradiol level in the Slavs compared to the Buryats (the Yakuts occupied an intermediate position). The established hormonal differences indicate ethnic features in the functioning of the hypothalamic-pituitary-testicular axis and suggest that they may have formed during the embryonic or pubertal periods, and led to ethnic differences in hormone-dependent anthropometric characteristics in adults. We found a negative correlation between estradiol levels and arm span in Buryats, a positive correlation between estradiol levels and leg length in Yakuts, and a positive correlation between testosterone levels and leg length in Slavs. Thus, it seems that in Buryats and Yakuts, estradiol acts as a determinant of the longitudinal growth of the skeleton, determining the stature, while in Slavs testosterone performs this function. Based on these facts, it can be assumed that in adolescence in Buryats and Yakuts, increased estradiol levels and reduced testosterone levels, unlike in Slavs, can contribute to earlier closure of the epiphyseal plates of tubular bones, thereby stopping their further growth and delaying the growth of the skeleton.

The length of AR CAG repeats was the shortest among the Slavs, higher among the Buryats and the highest among the

Table 3. Comparison of anthropometric and hormonal parameters according to the length of the AR CAG repeats in men of three ethnic groups

Buryats			
Parameter	Short CAG ≤ 20, n = 29	Medium 21 ≤ CAG ≤ 27, n = 166	Long CAG ≥ 28, n = 27
Height, cm	174.7 (7.2) ^a	175.4 (6.3) ^a	171.6 (4.9)^b
Leg length, cm	88.0 (5.2) ^a	89.7 (5.2) ^b	89.7 (4.0) ^a
TI	1.99 (0.06) ^a	1.96 (0.07) ^a	1.96 (0.08) ^a
Arm span, cm	178.0 (7.5) ^a	178.0 (7.5) ^a	176.6 (6.6) ^a
ADI	1.019 (0.026) ^{ab}	1.015 (0.024) ^a	1.029 (0.027)^b
2D-right, cm	7.3 (0.8) ^a	7.7 (1.1) ^a	7.6 (1.0) ^a
2D-left, cm	7.4 (0.8) ^a	7.8 (1.1) ^a	7.6 (1.0) ^a
4D-right, cm	7.9 (0.9) ^a	8.2 (1.3) ^a	8.0 (1.2) ^a
4D-left, cm	7.8 (0.9) ^a	8.2 (1.3) ^a	8.0 (1.1) ^a
2D:4D-right	0.94 (0.04) ^a	0.95 (0.03) ^{ab}	0.95 (0.02)^b
2D:4D-left	0.94 (0.04) ^a	0.95 (0.03) ^a	0.96 (0.02) ^a
Testosterone, nmol/L	17.78 (4.89) ^a	18.03 (6.02) ^a	19.31 (7.52) ^a
Estradiol, nmol/L	0.211 (0.036) ^a	0.226 (0.054) ^{ab}	0.248 (0.098)^b
Slavs			
Parameter	Short CAG ≤ 19, n = 60	Medium 20 ≤ CAG ≤ 24, n = 430	Long CAG ≥ 25, n = 212
Height, cm	179.3 (7.9) ^a	179.2 (6.6) ^a	179.1 (6.6) ^a
Leg length, cm	92.5 (5.7) ^a	94.1 (5.1) ^b	94.1 (5.5)^b
TI	1.89 (0.09) ^a	1.91 (0.08) ^a	1.91 (0.09) ^a
Arm span, cm	182.0 (8.7) ^a	182.3 (8.0) ^a	182.4 (8.0) ^a
ADI	1.015 (0.033) ^a	1.019 (0.030) ^a	1.019 (0.028) ^a
2D-right, cm	7.5 (0.7) ^a	7.8 (1.1) ^b	7.8 (1.0) ^{ab}
2D-left, cm	7.6 (0.7) ^a	7.8 (1.1) ^a	7.8 (1.0) ^a
4D-right, cm	7.8 (0.6) ^a	8.0 (1.1) ^a	7.9 (1.1) ^a
4D-left, cm	7.7 (0.8) ^a	8.0 (1.1) ^a	8.0 (1.1) ^a
2D:4D-right	0.97 (0.04) ^a	0.98 (0.04) ^a	0.98 (0.04) ^a
2D:4D-left	0.98 (0.04) ^a	0.98 (0.04) ^a	0.98 (0.04) ^a
Testosterone, nmol/L	19.43 (6.33) ^a	20.97 (7.41) ^{ab}	22.14 (8.13)^b
Estradiol, nmol/L	0.186 (0.056) ^a	0.197 (0.070) ^a	0.189 (0.062) ^a
Yakuts			
Parameter	Short CAG ≤ 22, n = 30	Medium 23 ≤ CAG ≤ 27, n = 96	Long CAG ≥ 28, n = 21
Height, cm	171.3 (8.1) ^a	172.5 (6.1) ^a	172.3 (6.3) ^a
Leg length, cm	88.2 (6.5) ^a	89.7 (5.1) ^a	89.9 (6.0) ^a
TI	1.95 (0.09) ^a	1.93 (0.08) ^a	1.92 (0.08) ^a
Arm span, cm	175.3 (9.9) ^a	178.3 (6.5) ^b	178.3 (7.1)^b
ADI	1.023 (0.022) ^a	1.034 (0.021) ^b	1.035 (0.020)^b
2D-right, cm	7.8 (1.1) ^a	7.7 (0.9) ^a	8.0 (1.0) ^a
2D-left, cm	7.8 (1.0) ^a	7.7 (0.9) ^a	8.0 (1.0) ^a
4D-right, cm	8.3 (1.1) ^a	8.1 (1.0) ^a	8.4 (1.2) ^a
4D-left, cm	8.3 (1.1) ^a	8.2 (1.0) ^a	8.4 (1.1) ^a
2D:4D-right	0.95 (0.03) ^a	0.95 (0.03) ^a	0.95 (0.04) ^a
2D:4D-left	0.94 (0.03) ^a	0.94 (0.04) ^a	0.94 (0.04) ^a
Testosterone, nmol/L	18.43 (5.48) ^a	19.58 (6.76) ^{ab}	21.97 (7.01)^b
Estradiol, nmol/L	0.227 (0.052) ^a	0.223 (0.056) ^a	0.221 (0.058) ^a

Yakuts. Positive correlations between the AR CAG repeat length and the estradiol levels in Buryats or the testosterone levels in Yakuts were supplemented by the association of long CAG repeats with increased estradiol levels in Buryats and increased testosterone levels in Yakuts and Slavs. Thus, it were the long CAG repeats that coordinated the variability of steroid hormone levels in all three ethnic groups. Taking into account the role of sex steroid hormones as important regulators of skeletal bone growth and bone tissue homeostasis, it can be assumed that sex steroid hormones can mediate the ethno-specific effects of AR CAG repeats on male anthropometric characteristics. Moreover, the AR CAG repeat polymorphism can predict the sex steroid level in men of the ethnic group studied.

In the current paper, the identified effects of long AR CAG repeats on the steroid hormone levels coincide with those previously obtained in European men and generally confirm that individual variability of testosterone and/or estradiol levels in men may be partially due to the AR CAG repeats polymorphism (Crabbe et al., 2007; Huhtaniemi et al., 2009; De Naeyer et al., 2014). In aging men from 8 European countries, AR CAG repeat length positively correlated with serum testosterone and estradiol levels, while higher testosterone levels in men with long AR CAG repeats corresponded to a lack of age-related hypogonadism in these patients (Huhtaniemi et al., 2009).

Hormonal effects of the AR CAG polymorphism have not been confirmed in Filipino men (Ryan et al., 2017) and Greek men (Goutou et al., 2009). The discrepancy in the results may be a consequence of ethno-specific characteristics or mixed ethnic composition of the studied groups. It should be noted that lifestyle factors (obesity, physical inactivity, taking anabolic steroids, stress, etc.) can affect the hormonal background, and altered levels of testosterone or estradiol will mask the genetic effects of the AR CAG polymorphism in men (Wrzosek et al., 2020).

Studies *in vitro* and *in vivo* demonstrated that the longer the length of AR CAG repeats, the weaker the transactivation ability of AR and the weaker the effects of androgens in target tissues. A possible mechanism underlying this phenomenon may be related to specific proteins known as coregulators that modulate the transcriptional activity of androgen-bound AR (Buchanan et al., 2004; Davey, Grossmann, 2016). In Slavs and Yakuts, an increase in the CAG repeat length is accompanied by an increase in testosterone levels, which compensates for a decrease in AR functional activity. In addition, in the hypothalamic-pituitary-testicular system, longer AR CAG repeats weaken androgen feedback and increase testosterone levels (Huhtaniemi et al., 2009). From a genetic point of view, the AR CAG polymorphism is one of many genetic factors underlying the genetic control of testosterone levels as a polygenic trait.

The normal AR function is maintained in a critical and limited range of CAG repeat lengths. Molecular modeling revealed a critical range of 16–29 triplets that would maintain maximum interaction between the transactivating domain and hormone binding domain of the AR (Nenonen et al., 2010). Consequently, the analysis of CAG repeat length in linear

regression models, performed in most studies, is probably not adequate enough, and data stratification may be an alternative way to study the relationship of CAG repeat length with phenotypic traits. Indeed, in our study, we failed to establish the linear relationship between the AR CAG repeat length and the steroid hormone level in the Slavic group, however, when stratifying CAG repeats into short, medium and long, the effects of long CAG repeats were revealed. A similar way has been successfully applied in other studies that have established the stimulating effects of long AR CAG repeats on a number of hormonal and anthropometric traits, including the level of total and/or free testosterone and/or estradiol (Crabbe et al., 2007; De Naeyer et al., 2014; Khan et al., 2018), as well as height, body weight, bone mineral density in adolescent boys (Rodríguez-García et al., 2015).

As already mentioned, the testosterone effect on bone growth and metabolism is exerted together with estradiol after the conversion of testosterone into estradiol by the aromatase enzyme, and is mediated by estrogen receptors ER α , ER β (Almeida et al., 2017; Alemany, 2022). Probably, in Buryat carriers of long AR CAG repeats, the estrogen effect on some anthropometric indicators (height, ADI, the 2D:4D ratio) may be due to the increased aromatase activity involved in the action of androgen and AR on bone growth.

The variability of the AR CAG repeat length associated with testosterone levels can already affect anthropometric parameters in embryogenesis, including finger length and the 2D:4D ratio (Manning et al., 2002; McIntyre, 2006; Grimbos et al., 2010; Knickmeyer et al., 2011; Folland et al., 2012). In our study, the CAG repeat length positively correlated with the 2D:4D ratio, and long AR CAG repeats increased the 2D:4D ratio in Buryats; it could probably be due to lower prenatal androgenization. However, we found a positive relationship between testosterone levels and both 2D:4D ratios in Slavic men, which does not seem to be associated with long AR CAG repeats. It is worth noting that the validity of the 2D:4D ratio remains controversial, since the research data are very contradictory. There are studies both confirming and not confirming the relationship of the digit index with the level of testosterone and/or estradiol or with the CAG repeat length in adult men (Hönekopp et al., 2007; Knickmeyer et al., 2011; Muller et al., 2011; Hönekopp, 2013; Zhang et al., 2013). In animal studies, there is more reliable experimental evidence of the effect of prenatal sex hormones on the 2D:4D ratio and its relation to the AR CAG repeat length (Zheng, Cohn, 2011; Swift-Gallant et al., 2020). Our results complement the existing studies, but indicate the ethnic characteristics of such associations.

Conclusion

The study found 1) differences in hormonal and anthropometric indicators between men of three ethnic groups: Buryats, Yakuts and Slavs; 2) ethno-specific correlations between hormonal and anthropometric indicators; 3) ethno-specific positive correlations between the AR CAG repeat length and sex steroid levels; 4) ethno-specific effects of long AR CAG repeats on selected anthropometric traits and sex steroid levels in all ethnic groups.

References

- Alemamy M. The roles of androgens in humans: Biology, metabolic regulation and health. *Int. J. Mol. Sci.* 2022;23(19):11952. DOI 10.3390/ijms231911952
- Almeida M., Laurent M.R., Dubois V., Claessens F., O'Brien C.A., Bouillon R., Vanderschueren D., Manolagas S.C. Estrogens and androgens in skeletal physiology and pathophysiology. *Physiol. Rev.* 2017;97(1):135-187. DOI 10.1152/physrev.00033.2015
- Buchanan G., Yang M., Cheong A., Harris J.M., Irvine R.A., Lambert P.F., Moore N.L., Raynor M., Neufing P.J., Coetzee G.A., Tilley W.D. Structural and functional consequences of glutamine tract variation in the androgen receptor. *Hum. Mol. Genet.* 2004;13(16):1677-1692. DOI 10.1093/hmg/ddh181
- Crabbe P., Bogaert V., De Bacquer D., Goemaere S., Zmierzczak H., Kaufman J.M. Part of the interindividual variation in serum testosterone levels in healthy men reflects differences in androgen sensitivity and feedback set point: contribution of the androgen receptor polyglutamine tract polymorphism. *J. Clin. Endocrinol. Metab.* 2007;92(9):3604-3610. DOI 10.1210/jc.2007-0117
- Davey R.A., Grossmann M. Androgen receptor structure, function and biology: From bench to bedside. *Clin. Biochem. Rev.* 2016;37(1):3-15
- Davis-Dao C.A., Tuazon E.D., Sokol R.Z., Cortessis V.K. Male infertility and variation in CAG repeat length in the androgen receptor gene: a meta-analysis. *J. Clin. Endocrinol. Metab.* 2007;92(11):4319-4326. DOI 10.1210/jc.2007-1110
- De Naeyer H., Bogaert V., De Spaey A., Roef G., Vandewalle S., Derave W., Taes Y., Kaufman J.M. Genetic variations in the androgen receptor are associated with steroid concentrations and anthropometrics but not with muscle mass in healthy young men. *PLoS One.* 2014;9(1):e86235. DOI 10.1371/journal.pone.0086235
- Folland J.P., Mc Cauley T.M., Phipers C., Hanson B., Mastana S.S. Relationship of 2D:4D finger ratio with muscle strength, testosterone, and androgen receptor CAG repeat genotype. *Am. J. Phys. Anthropol.* 2012;148(1):81-87. DOI 10.1002/ajpa.22044
- Frank G.R. Role of estrogen and androgen in pubertal skeletal physiology. *Med. Pediatr. Oncol.* 2003;41(3):217-221. DOI 10.1002/mpo.10340
- Goutou M., Sakka C., Stakias N., Stefanidis I., Koukoulis G.N. AR CAG repeat length is not associated with serum gonadal steroids and lipid levels in healthy men. *Int. J. Androl.* 2009;32(6):616-622. DOI 10.1111/j.1365-2605.2008.00908.x
- Grigорова M., Punab M., Kahre T., Ivandi M., Tönissон N., Poolamets O., Vihlajev V., Žilaitienė B., Erenpreiss J., Matulevičius V., Laan M. The number of CAG and GGN triplet repeats in the androgen receptor gene exert combinatorial effect on hormonal and sperm parameters in young men. *Andrology.* 2017;5(3):495-504. DOI 10.1111/andr.12344
- Grimbos T., Dawood K., Burriss R.P., Zucker K.J., Puts D.A. Sexual orientation and the second to fourth finger length ratio: A meta-analysis in men and women. *Beh. Neurosci.* 2010;124(2):278-287. DOI 10.1037/a0018764
- Guadalupe-Grau A., Rodríguez-González F.G., Ponce-González J.G., Dorado C., Olmedillas H., Fuentes T., Pérez-Gómez J., Sanchis-Moys J., Díaz-Chico B.N., Calbet J.A. Bone mass and the CAG and GGN androgen receptor polymorphisms in young men. *PLoS One.* 2010;5(7):e11529. DOI 10.1371/journal.pone.0011529
- Hönekopp J. No evidence that 2D:4D is related to the number of CAG repeats in the androgen receptor gene. *Front. Endocrinol. (Lausanne).* 2013;4:185. DOI 10.3389/fendo.2013.00185
- Hönekopp J., Bartholdt L., Beier L., Liebert A. Second to fourth digit length ratio (2D:4D) and adult sex hormone levels: new data and a meta-analytic review. *Psychoneuroendocrinology.* 2007;32(4):313-321. DOI 10.1016/j.psyneuen.2007.01.007
- Huhtaniemi I.T., Pye S.R., Limer K.L., Thomson W., O'Neill T.W., Platt H., Payne D., John S.L., Jiang M., Boonen S., Borghs H., Vanderschueren D., Adams J.E., Ward K.A., Bartfai G., Casanueva F., Finn J.D., Forti G., Giwercman A., Han T.S., Kula K., Lean M.E., Pendleton N., Punab M., Silman A.J., Wu F.C., European Male Ageing Study Group. Increased estrogen rather than decreased androgen action is associated with longer androgen receptor CAG repeats. *J. Clin. Endocrinol. Metab.* 2009;94(1):277-284. DOI 10.1210/jc.2008-0848
- Jägetoft Z., Unenge Hallerbäck M., Julin M., Bornehag C.G., Wikström S. Anthropometric measures do not explain the 2D:4D ratio sexual dimorphism in 7-year-old children. *Am. J. Hum. Biol.* 2022;34(9):e23776. DOI 10.1002/ajhb.23776
- Khan H.L., Bhatti S., Abbas S., Khan Y.L., Aslamkhan M., Gonzalez R.M.M., Gonzalez G.R., Aydin H.H., Trinidad M.S. Tri-nucleotide consortium of androgen receptor is associated with low serum FSH and testosterone in asthenospermic men. *Syst. Biol. Reprod. Med.* 2018;64(2):112-121. DOI 10.1080/19396368.2017.1384080
- Knickmeyer R.C., Woolson S., Hamer R.M., Konneker T., Gilmore J.H. 2D:4D ratios in the first 2 years of life: Stability and relation to testosterone exposure and sensitivity. *Horm. Behav.* 2011;60(3):256-263. DOI 10.1016/j.yhbeh.2011.05.009
- Ma Y.M., Wu K.J., Ning L., Zeng J., Kou B., Xie H.J., Ma Z.K., Wang X.Y., Gong Y.G., He D.L. Relationships among androgen receptor CAG repeat polymorphism, sex hormones and penile length in Han adult men from China: a cross-sectional study. *Asian J. Androl.* 2014;16(3):478-481. DOI 10.4103/1008-682X.124560
- Manning J., Kilduff L., Cook C., Crewther B., Fink B. Digit Ratio (2D:4D): A biomarker for prenatal sex steroids and adult sex steroids in challenge situations. *Front. Endocrinol. (Lausanne).* 2014;5:9. DOI 10.3389/fendo.2014.00009
- McIntyre M.H. The use of digit ratios as markers for perinatal androgen action. *Reprod. Biol. Endocrinol.* 2006;4:10. DOI 10.1186/1477-7827-4-10
- Mitsui T., Araki A., Imai A., Sato S., Miyashita C., Ito S., Sasaki S., Kitta T., Moriya K., Cho K., Morioka K., Kishi R., Nonomura K. Effects of prenatal Leydig cell function on the ratio of the second to fourth digit lengths in school-aged children. *PLoS One.* 2015;10(3):e0120636. DOI 10.1371/journal.pone.0120636
- Muller D.C., Giles G.G., Bassett J., Morris H.A., Manning J.T., Hopper J.L., English D.R., Severi G. Second to fourth digit ratio (2D:4D) and concentrations of circulating sex hormones in adulthood. *Reprod. Biol. Endocrinol.* 2011;9:57. DOI 10.1186/1477-7827-9-57
- Nenonen H., Björk C., Skjaerpe P.A., Giwercman A., Rylander L., Svartberg J., Giwercman Y.L. CAG repeat number is not inversely associated with androgen receptor activity in vitro. *Mol. Hum. Reprod.* 2010;16(3):153-157. DOI 10.1093/molehr/gap097
- Osadchuk L.V., Osadchuk A.V. Role of CAG and GGC polymorphism in the androgen receptor gene in male fertility. *Russ. J. Genet.* 2022;58(3):247-264. DOI 10.1134/S1022795422020119
- Osadchuk L., Vasiliev G., Kleshchev M., Osadchuk A. Androgen receptor gene CAG repeat length varies and affects semen quality in an ethnic-specific fashion in young men from Russia. *Int. J. Mol. Sci.* 2022;23(18):10594. DOI 10.3390/ijms231810594
- Richards G., Browne W.V., Aydin E., Constantinescu M., Nave G., Kim M.S., Watson S.J. Digit ratio (2D:4D) and congenital adrenal hyperplasia (CAH): Systematic literature review and meta-analysis. *Horm. Behav.* 2020;126:104867. DOI 10.1016/j.yhbeh.2020.104867
- Rodríguez-García L., Ponce-González J.G., González-Henríquez J.J., Rodríguez-González F.G., Díaz-Chico B.N., Calbet J.A., Serrano-Sánchez J.A., Dorado C., Guadalupe-Grau A. Androgen receptor CAG and GGN repeat polymorphisms and bone mass in boys and girls. *Nutr. Hosp.* 2015;32(6):2633-2639. DOI 10.3305/nh.2015.32.6.9767
- Ryan C.P., McDade T.W., Gettler L.T., Eisenberg D.T., Rzhetskaya M., Hayes M.G., Kuzawa C.W. Androgen receptor CAG repeat polymorphism and hypothalamic-pituitary-gonadal function in Filipino young adult males. *Am. J. Hum. Biol.* 2017;29(1):e22897. DOI 10.1002/ajhb.22897

- Swift-Gallant A., Johnson B.A., Di Rita V., Breedlove S.M. Through a glass, darkly: Human digit ratios reflect prenatal androgens, imperfectly. *Horm. Beh.* 2020;120:104686. DOI 10.1016/j.yhbeh.2020.104686
- Voorhoeve P.G., van Mechelen W., Uitterlinden A.G., Delemarre-van de Waal H.A., Lamberts S.W. Androgen receptor gene CAG repeat polymorphism in longitudinal height and body composition in children and adolescents. *Clin. Endocrinol.* 2011;74(6):732-735. DOI 10.1111/j.1365-2265.2011.03986.x
- Wang Y., Wei Y., Tang X., Liu B., Shen L., Long C., Lin T., He D., Wu S., Wei G. Association between androgen receptor polymorphic CAG and GGC repeat lengths and cryptorchidism: A meta-analysis of case-control studies. *J. Pediatr. Urol.* 2018;14(5):432.e1-432.e9. DOI 10.1016/j.jpuro.2018.05.011
- Wrzosek M., Woźniak J., Włodarek D. The causes of adverse changes of testosterone levels in men. *Expert. Rev. Endocrinol. Metab.* 2020; 15(5):355-362. DOI 10.1080/17446651.2020.1813020
- Xiao F., Lan A., Lin Z., Song J., Zhang Y., Li J., Gu K., Lv B., Zhao D., Zeng S., Zhang R., Zhao W., Pan Z., Deng X., Yang X. Impact of CAG repeat length in the androgen receptor gene on male infertility – a meta-analysis. *Reprod. Biomed. Online.* 2016;33(1):39-49. DOI 10.1016/j.rbmo.2016.03.012
- Xu Y., Zheng Y. The digit ratio (2D:4D) in China: A meta-analysis. *Am. J. Hum. Biol.* 2015;27(3):304-309. DOI 10.1002/ajhb.22639
- Zhang C., Dang J., Pei L., Guo M., Zhu H., Qu L., Jia F., Lu H., Huo Z. Relationship of 2D:4D finger ratio with androgen receptor CAG and GGN repeat polymorphism. *Am. J. Hum. Biol.* 2013;25(1):101-106. DOI 10.1002/ajhb.22347
- Zheng Z., Cohn M.J. Developmental basis of sexually dimorphic digit ratios. *Proc. Natl. Acad. Sci. USA.* 2011;108(39):16289-16294. DOI 10.1073/pnas.1108312108
- Zitzmann M., Nieschlag E. The CAG repeat polymorphism within the androgen receptor gene and maleness. *Int. J. Androl.* 2003;26(2): 76-83. DOI 10.1046/j.1365-2605.2003.00393.x

ORCID

L.V. Osadchuk orcid.org/0000-0002-7597-9204
G.V. Vasiliev orcid.org/0000-0003-0929-6832
A.V. Osadchuk orcid.org/0000-0002-4210-7354

Acknowledgements. This study was supported by the Russian Science Foundation (grant No. 19-15-00075).

Conflict of interest. The authors declare no conflict of interest.

Received May 5, 2023. Revised June 13, 2023. Accepted June 14, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Genetic history of the Koryaks and Evens of the Magadan region based on Y chromosome polymorphism data

B.A. Malyarchuk , M.V. Derenko

Institute of Biological Problems of the North of the Far-Eastern Branch of the Russian Academy of Sciences, Magadan, Russia
 malyarchuk@ibpn.ru

Abstract. In order to clarify the history of gene pool formation of the indigenous populations of the Northern Priokhotye (the northern coast of the Sea of Okhotsk), Y-chromosome polymorphisms were studied in the Koryaks and Evens living in the Magadan region. The results of the study showed that the male gene pool of the Koryaks is represented by haplogroups C-B90-B91, N-B202, and Q-B143, which are also widespread in other peoples of Northeastern Siberia, mainly of Paleo-Asiatic origin. High frequency of haplogroup C-B80, typical of other Tungus-Manchurian peoples, is characteristic of the Evens of the Magadan region. The shared components of the gene pools of the Koryaks and Evens are haplogroups R-M17 and I-P37.2 inherited as a result of admixture with Eastern Europeans (mainly Russians). The high frequency of such Y chromosome haplogroups in the Koryaks (16.7 %) and Evens (37.8 %) is indicative of close interethnic contacts during the last centuries, and most probably especially during the Soviet period. The genetic contribution of the European males' Y chromosome significantly prevails over that of maternally inherited mitochondrial DNA. The study of the Y chromosome haplogroup diversity has shown that only relatively young phylogenetic branches have been preserved in the Koryak gene pool. The age of the oldest component of the Koryak gene pool (haplogroup C-B90-B91) is estimated to be about 3.8 thousand years, the age of the younger haplogroups Q-B143 and N-B202 is about 2.8 and 2.4 thousand years, respectively. Haplogroups C-B90-B91 and N-B202 are Siberian in origin, and haplogroup Q-B143 was apparently inherited by the ancestors of the Koryaks and other Paleo-Asiatic peoples from the Paleo-Eskimos as a result of their migrations to Northeast Asia from the Americas. The analysis of microsatellite loci for haplogroup Q-B143 in the Eskimos of Greenland, Canada and Alaska as well as in the indigenous peoples of Northeastern Siberia showed a decrease in genetic diversity from east to west, pointing to the direction of distribution of the Paleo-Eskimo genetic component in the circumpolar region of America and Asia. At the same time, the Evens appeared in the Northern Priokhotye much later (in the XVII century) as a result of the expansion of the Tungusic tribes, which is confirmed by the results of the analysis of haplogroup C-B80 polymorphisms.

Key words: Y chromosome; polymorphism; human populations; the Koryaks; the Evens; genetic history.

For citation: Malyarchuk B.A., Derenko M.V. Genetic history of the Koryaks and Evens of the Magadan region based on Y chromosome polymorphism data. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(1):90-97. DOI 10.18699/vjgb-24-11

Генетическая история коряков и эвенов Магаданской области по данным о полиморфизме Y-хромосомы

Б.А. Мальярчук , М.В. Деренко

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия
 malyarchuk@ibpn.ru

Аннотация. Для прояснения истории формирования генофондов коренного населения Северного Приохотья изучен полиморфизм Y-хромосомы у коряков и эвенов Магаданской области. Результаты исследования показали, что мужской генофонд коряков представлен гаплогруппами C-B90-B91, N-B202, Q-B143, распространенными также у других народов Северо-Востока Сибири преимущественно палеоазиатского происхождения. Для эвенов Магаданской области характерна высокая частота гаплогруппы C-B80, свойственной и для других тунгусо-маньчжурских народов. Общим компонентом генофондов коряков и эвенов являются гаплогруппы R-M17 и I-P37.2, унаследованные в результате метисации с пришлым восточноевропейским (преимущественно русским) населением. Высокая частота такого рода гаплогрупп Y-хромосомы у коряков (16.7 %) и эвенов (37.8 %) свидетельствует об интенсивных межэтнических контактах на протяжении последних столетий, и особенно, по всей видимости, в советское время. Причем генетический вклад со стороны европейских мужчин (по Y-хромосоме) существенно преобладает над таковым со стороны женщин (по митохондриальной ДНК). Исследование разнообразия гаплогрупп Y-хромосомы показало, что в генофонде

коряков сохранились только относительно молодые филогенетические ветви. Возраст наиболее древнего компонента генофонда коряков (гаплогруппа С-В90-В91) оценивается примерно в 3.8 тыс. лет, возраст более молодых гаплогрупп Q-В143 и N-В202 составляет примерно 2.8 и 2.4 тыс. лет соответственно. Гаплогруппы С-В90-В91 и N-В202 являются сибирскими по происхождению, а гаплогруппа Q-В143, вероятно, унаследована предками коряков (и других палеоазиатских народов) от палеоэскимосов в результате их миграций на Северо-Восток Азии из Америки. Анализ микросателлитных локусов для гаплогруппы Q-В143 у эскимосов Гренландии, Канады и Аляски, а также у представителей коренного населения Северо-Востока Сибири выявил снижение генетического разнообразия с востока на запад, что указывает на направление распространения палеоэскимосского генетического компонента в циркумполярном регионе Америки и Азии. Эвены же появились в Северном Приохотье намного позже (в XVII в.) в результате экспансии тунгусских племен, что подтверждается данными анализа полиморфизма гаплогруппы С-В80.

Ключевые слова: Y-хромосома; полиморфизм; популяции человека; коряки; эвены; генетическая история.

Introduction

The extreme Northeast of Siberia is inhabited by the Chukotka-Kamchatkan peoples (the Chukchis, Koryaks, Itelmens) and the Eskimos, which are characterized by genetic peculiarities and occupy a distinct position among the ethnogeographical groups of Northern Eurasia (Rasmussen et al., 2010; Fedorova et al., 2013; Cardona et al., 2014; Pagani et al., 2016; Pugach et al., 2016; Gorin et al., 2022). According to paleogenomic data, the genetic specificity of these peoples is due to their ancient Paleo-Siberian genetic substrate, inherited in part by the Native Americans (Sikora et al., 2019). Meanwhile, the results of the analysis of autosomal loci polymorphism in the indigenous Siberian populations have shown that in the east of Siberia the appearance of alleles of European origin is estimated to be relatively recent (about 3–6 generations ago), which is associated with the Russian discovery of Siberia, starting mainly from the XVII century and especially intensive during the Soviet period (Cardona et al., 2014). Moreover, various studies demonstrate that the flow of European genes into the gene pools of the indigenous populations of Northeastern Siberia was carried out predominantly by men (Balanovska et al., 2020a, b; Agdzhoyan et al., 2021; Solovyev et al., 2023). In this regard, the contribution of European Y chromosome variants to the gene pools of indigenous peoples of Northeastern Siberia and other Arctic regions usually exceeds that of European maternally inherited mitochondrial DNA (mtDNA) variants (Bosch et al., 2003; Rubicz et al., 2010; Dulik et al., 2012; Olofsson et al., 2015).

The results of genetic studies of the indigenous populations of the northern coast of the Sea of Okhotsk – the Koryaks and Evens of the Magadan region – have shown that they have a very low frequency of European mtDNA variants (only in the Evens it reaches 4 %) (Derenko et al., 2023), and according to the results of genome-wide analysis, the frequency of the European genetic component in the Northeastern Siberian populations has significantly increased only in the last ~100 years (Cardona et al., 2014). Most likely, this may be related to the increased European contribution by males, and therefore, the aim of this paper is to analyze the Y chromosome polymorphism in the indigenous populations of the Magadan region.

Materials and methods

Unrelated males from the indigenous populations of the Magadan region (the Koryaks and Evens) were studied (Sup-

plementary Materials 1 and 2)¹. Based on survey data, the Koryaks ($N = 36$) and Evens ($N = 61$) studied had identified themselves as belonging to the above ethnic groups for at least 2–3 generations. According to the results of mtDNA analysis, all individuals studied are characterized by haplotypes of Northeast Asian origin.

DNA was extracted and purified from whole blood as we previously described (Derenko, Malyarchuk, 2010). Samples were genotyped for 12 microsatellite (STR) loci (DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439) using PowerPlex Y System (Promega Corporation, Madison, WI, USA). Alleles were detected by capillary electrophoresis on ABI 3500xL Genetic Analyzer (Applied Biosystems, USA). The results were analyzed using the programs Genscan v. 3.7 and Genotyper v. 3.7 (Applied Biosystems). Data for DYS385 loci were not considered in the statistical analysis because the order of the DYS385a and DYS385b loci on the Y chromosome is unknown. The number of repeats at the DYS389II locus was determined by subtracting the length of the smaller repeat (DYS389I) from the length of the larger repeat (DYS389II).

Y chromosome haplogroups were determined by direct DNA sequencing or restriction fragment length polymorphism analysis of haplogroup markers as we described previously (Malyarchuk et al., 2013). Data on variability of the B77, B79, B80, B81, B90, B91, B92, B94, B143, B186, B202, B203, B204, and B471 loci were obtained earlier in studies of whole Y chromosome variability in different ethnic groups, including some Koryak and Even individuals from the Magadan region (Karmin et al., 2015).

The V_p statistic, the average dispersion of the number of repeats in STR loci, was used to estimate intrapopulation genetic diversity (Kayser et al., 2001). The evolutionary age of the Y chromosome haplogroups was calculated based on the analysis of the average number of repeats in loci and their variance (Zhitovovsky et al., 2004). The mutation rate value used in the calculations, $2.79 \cdot 10^{-3}$ substitutions per locus per generation, was obtained by averaging mutation rates for the 10 Y chromosome loci analyzed, according to Ballantyne et al. (2010). The program Network 10.2 (www.fluxus-engineering.com) was used to construct median networks of the Y chromosome STR haplotypes.

¹ Supplementary Materials 1 and 2 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Malyar_Engl_28_1.pdf

Results and discussion

The results of the study of Y chromosome polymorphism showed that the male gene pool of the Koryaks living in the Magadan region is represented mainly by haplogroups C, N, and Q (Table 1). European lineages in the Koryaks were found at a frequency of 16.7 % for haplogroups R-M17, I-M253, and I-P37.2. The frequency of European haplogroups is even higher among the Evens – 37.8 %. They are represented by haplogroups R-M17, R-M269, I-P37.2, as well as N-B186, which is characteristic of the peoples of Northeastern Europe (Karmin et al., 2015). The East Asian component of the Even gene pool consists of various subgroups of haplogroup C (55.7 % in total). In addition, haplotypes belonging to haplogroup Q-M3, which is widespread among the Native Americans and Eskimos, have been found in the Evens.

Haplogroup C variants in the Koryak and Even populations differ significantly. The Koryaks are characterized by the B90 and B91 specific markers, while the Evens fall into the B80-defined subgroup. According to the results of whole-genome studies, the B90 marker is specific for the Y chromosomes of the indigenous populations of Northeastern Siberia (the Koryaks, Evenks, and Ulchi) (Karmin et al., 2015; Balanovska et al., 2018), and the B91-defined subgroup is present only in the Koryaks (Karmin et al., 2015). Its frequency in the Koryaks of the Magadan region is 27.8 % (see Table 1).

According to the results of molecular dating based on the analysis of polymorphism associated with single nucleotide substitutions (SNP) in whole Y chromosomes, the age of the B91 subgroup is estimated at 3.8 (3.0–4.7) thousand years (Karmin et al., 2015). The age of the upstream C-B90 subgroup is approximately 5.0 (4.2–5.7) thousand years. Based on the similarity of STR profiles, B90 haplotypes appear to be predominantly distributed in Northeastern Siberia, since, in addition to the Koryaks, Evenks and Ulchi, homologous STR haplotypes are observed in the Yakuts, Yukaghirs, Itelmens, and Evenks². In our study, a single homologous B90 haplotype (similar to that of the Koryaks) was also found in the Evens.

In the Evens, the C subgroup, marked by a substitution at locus B80, is mainly distributed (see Table 1). It is known that B80 haplotypes, in addition to the Evens, are also characteristic of other Tungus-Manchurian peoples (the Orochens, Evenks, and Manchurians) (Yu et al., 2023). The evolutionary age of this subgroup, according to the SNP data, is 1.7 (1.2–2.2) thousand years (Karmin et al., 2015). The results of the analysis by H.-X. Yu et al. (2023) have shown that the age of the B80 subgroup is estimated to be about 2 thousand years, while the B81 and B471 haplotypes specific to the Evens originated in the Amur region and spread to Northeastern Siberia as a result of the migrations of the Tungus ancestors in the last approximately 1.5 thousand years.

The N haplogroup in the Koryaks of the Magadan region is represented exclusively by the N-B202 branch (25 %). The same subgroup predominates in the gene pool of the Chukchi (Karmin et al., 2015; Ilumäe et al., 2016; Agdzhoyan et al., 2021), and is also found in the neighboring peoples – the Itelmens and Eskimos (Agdzhoyan et al., 2021). The age of

Table 1. Frequency (in %) of Y chromosome haplogroups in the Koryaks and Evens of the Magadan region

Haplogroup	Koryaks (N = 36)	Evens (N = 61)
C-M217-M48-B90-B91-B92	13.9	0
C-M217-M48-B90-B91-B94	13.9	0
C-M217-M48-B90	0	1.6
C-M217-M48-B80-B81	0	16.4
C-M217-M48-B80-B471	0	26.2
C-M217-M48	2.8	11.5
C-M217-B77	2.8	0
C-M217-B79	2.8	0
C-M217	2.8	0
N-B202-B203	5.6	0
N-B202-B204	19.4	0
N-B186	0	3.3
N-M46	0	3.3
Q-B143	16.7	0
Q-M346-M3	0	3.3
O-M122	2.8	0
R-M17	8.3	27.9
R-M269	0	3.3
I-M253	2.8	0
I-P37.2	2.8	3.3
J-M314	2.8	0

the N-B202 branch is approximately 2.4 (1.8–3.1) thousand years (Ilumäe et al., 2016). This haplogroup consists of two subgroups, the older N-B204 (estimated to be about 1.4 thousand years old based on STR haplotype diversity) and the younger N-B203 (about 600 years old) (Agdzhoyan et al., 2021). In the Chukchi, both subgroups are present to a nearly equal extent, with the older subgroup N-B204 predominating in the Koryaks (see Supplementary Material 1). In the Evens of the Magadan region, haplogroup N was found at a relatively low frequency (6.6 %) and is represented by different haplotypes. In this respect, the Magadan Evens are similar to the Kamchatkan Evens, but differ from the Okhotsk Evens, who are characterized by the “Amur region” subgroup N-B479 at a frequency of 10 % (Agdzhoyan et al., 2019).

Haplogroup Q represents the oldest component of the gene pools of the indigenous populations of Siberia and America. Haplogroup Q-F903 was found in an Upper Paleolithic inhabitant of Eastern Siberia (the Afontova Gora archaeological site, approximately 17 thousand years old) (Raghavan et al., 2014), and haplogroup Q-B143 was revealed in Northeastern Siberia (the Duvanniy Yar site, about 10 thousand years old) (Sikora et al., 2019). The same haplogroup was reported in

² Adamov D.S. Summary table of Y-STR haplotypes of haplogroup C-M48 of Yakut-Sakha. 2019. <https://www.researchgate.net/profile/Dmitry-Adamov/publications/> (Reference date: September 5, 2023).

Table 2. Diversity and evolutionary age of the Q-B143 STR haplotypes in the Eskimo and Paleo-Asiatic peoples

Geographic region	<i>N</i>	<i>n</i>	Vp	Age, thousand years
Northeast Asia	12	6	0.058	0.747 ± 0.401
Alaska	34	18	0.181	1.842 ± 0.525
Canada	28	20	0.160	1.703 ± 0.769
Greenland	70	27	0.208	2.456 ± 1.188

Note. *N* is the sample size, *n* is the number of STR haplotypes, Vp is the variance of the number of repeats in STR loci.

a representative of the Paleo-Eskimo Sakkak culture who lived in Greenland about 4 thousand years ago (Rasmussen et al., 2010). Currently, haplogroup Q-B143 is distributed only among the indigenous populations of the American Far North, Greenland and Siberia (Malyarchuk et al., 2011; Karmin et al., 2015; Grugni et al., 2019; Luis et al., 2023). In the Koryaks of the Magadan region, this Q haplogroup was detected with a frequency of 16.7 % (see Table 1). According to indirect data (based on the frequencies of haplogroups Q(xM346) and Q-NWT01, as well as on the similarity of STR haplotypes), haplogroup Q-B143 is present in the Koryaks of Kamchatka (with frequency varying from 6 to 18 %) (Karafet et al., 2018), in the Chukchi (13 %) (Pakendorf et al., 2006), and it has also been found with high frequencies (up to 50 %) in the Eskimos of Alaska, Canada and Greenland (Dulik et al., 2012; Olofsson et al., 2015; Luis et al., 2023).

The presence of haplogroup Q-B143 in the Northeast of Siberia about 10 thousand years ago and at present suggests that Q-B143 is the most ancient Siberian component that has been a part of the gene pools of the Paleo-Asiatic peoples and their ancestors. Archaeological data, as well as the results of the study of haplogroup Q polymorphism, showed that about 5 thousand years ago the carriers of haplogroup Q-B143 (as well as unsuccessful Q-L713 and Q-preM120) migrated from Siberia to America and then to Greenland and became the founders of the Paleo-Eskimo culture (Grugni et al., 2019). However, the results of the Q-B143 dating showed that the age of this haplogroup in modern Koryaks is only about 2.8 thousand years, which indicates the possibility of back migration of the carriers of these haplotypes (most likely, the Paleo-Eskimos) from North America to Northeast Asia (Grugni et al., 2019). Similarly, the results of studies of STR variability within haplogroup Q-B143 in Greenlandic and North American Eskimos showed that the diversity and evolutionary age of haplotypes in Greenlandic Eskimos are higher than in Canadian and Alaskan Eskimos (Olofsson et al., 2015; Luis et al., 2023). In this connection, these authors suggested that haplogroup Q-B143 was spread by the Paleo-Eskimos from the east to the west of America and, moreover, became one of the main components of the gene pool of the Neo-Eskimos, which most likely formed in the north of America about 700 years ago.

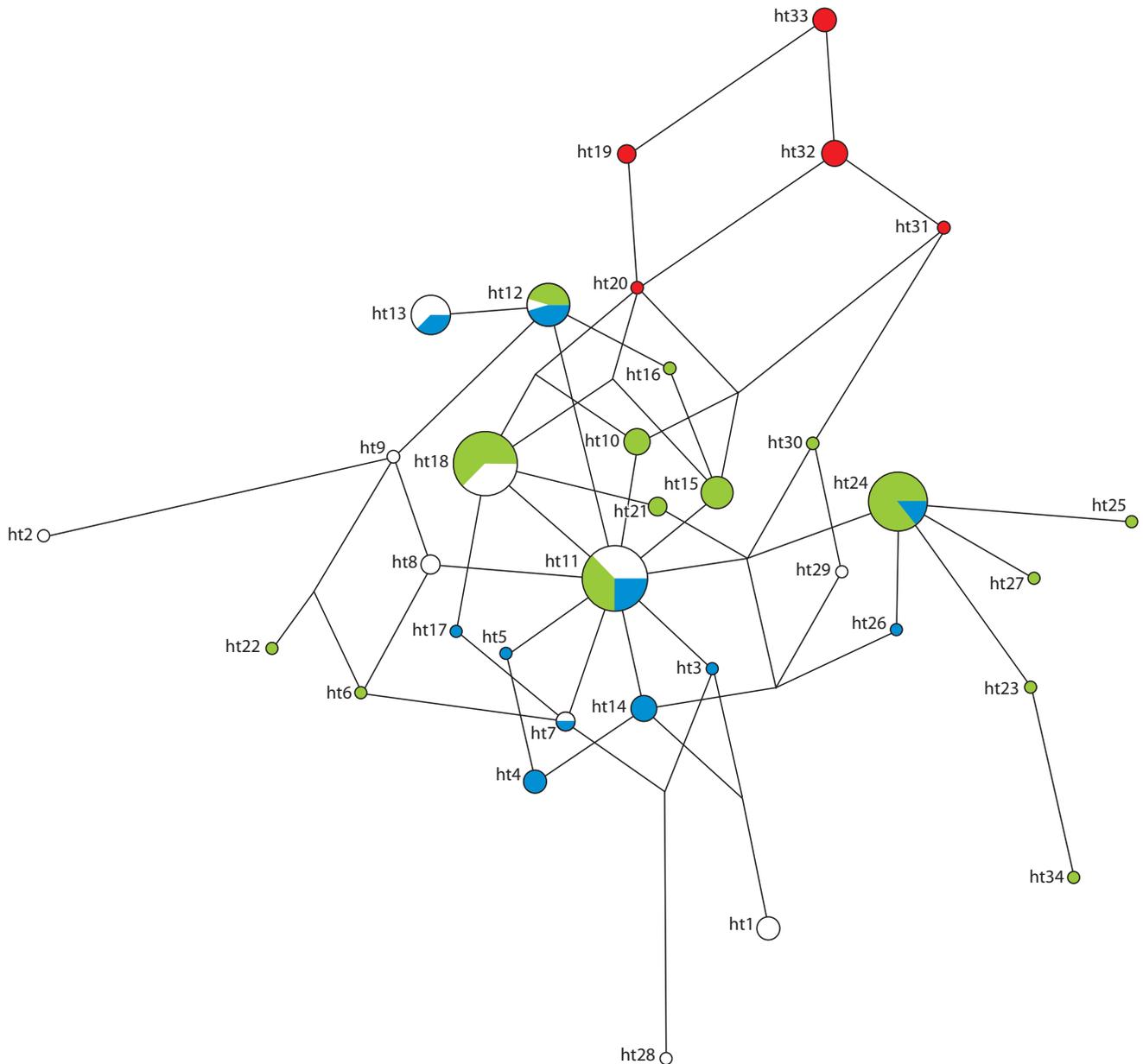
Since Luis et al. (2023) did not investigate Q-B143 haplotypes in the indigenous populations of Northeast Asia, we

analyzed STR haplotype diversity in samples of Greenlandic, Canadian, and Alaskan Eskimos (based on data from Dulik et al. (2012), Olofsson et al. (2015), Luis et al. (2023)), and in the Koryaks, Yukaghirs, and Chukotkan Eskimos (according to Pakendorf et al. (2006), Luis et al. (2023), and the present study). The results of our study showed that, indeed, Northeast Asian sample has the lowest diversity of Q-B143 haplotypes compared to Greenlandic and North American ones, indicating that these haplotypes appeared in Northeast Asia later than in North America and Greenland (Table 2).

It is necessary to note the discrepancy between the dates obtained using STR markers and whole-genome SNP data, because the evolutionary age of haplogroup Q-B143 in the Koryaks according to SNP data (2.8 ± 0.9 thousand years as per Grugni et al. (2019)) exceeds that obtained using STR markers for the indigenous population of Northeastern Siberia (0.7 ± 0.4 thousand years) (see Table 2). This is most likely due to the very large mismatch in the number of variable positions for the compared genetic systems, the high probability of recurrent (forward and reverse) mutations for rapidly evolving STR loci, and the dependence of such mutational events on the age of haplogroups. Therefore, it is likely that STR dates close to the whole-genome ones can be obtained only for young branches (Agdzhoyan et al., 2021). Thus, if we focus on the whole-genome SNP dating (as more accurate), we can assume that the appearance of haplogroup Q-B143 in Northeast Asia occurred long before the appearance of the Neo-Eskimos and is thus associated with the migrations of the Paleo-Eskimos. The possibility of such events is evidenced by archaeological data, according to which the Paleo-Eskimo cultural tradition was established in Chukotka about 3.0–3.5 thousand years ago (the Chertov Ovrag site on Wrangel Island and the Unenen settlement), as well as in the Sea of Okhotsk's northern coasts by representatives of the Tokarev culture (probable ancestors of the Koryaks) about 2.8 thousand years ago (Grebnyuk et al., 2019). The low level of diversity of Northeastern Siberian STR haplotypes and their peripheral position in the median network among the huge number of Q-B143 haplotypes of Arctic peoples indicate a very small number of successful (in terms of reproduction) migrations of the Paleo-Eskimos to the Asian coast (see the Figure). In fact, a single haplotype (ht20 in the Figure) is the most likely ancestor for the other haplotypes identified in the Koryaks and Yukaghirs.

The low level of heterogeneity of Q-B143 haplotypes in the indigenous populations of Northeastern Siberia also indicates that the most ancient haplotypes, ancestral to the haplotypes of the Paleo-Eskimos of the north of America and Greenland,

³ Kharkov V.N. Structure and phylogeography of the gene pool of the indigenous population of Siberia according to Y chromosome markers. Dr. Biol. Sci. Diss. Tomsk, 2012. (in Russian).



Median network of STR haplotypes belonging to Y chromosome haplogroup Q-B143 in the Eskimos of Greenland (green), Canada (blue), Alaska (white) and in the indigenous population of Northeast Siberia (the Koryaks, Yukaghirs, Eskimos) (red).

have not been preserved in their gene pools. This seems quite likely, given the low effective population size of Northeastern Siberians and the increasing role of genetic drift under these conditions, as well as the continuing influence from neighboring Siberian populations. It is known that periods of almost complete population replacements occurred more than once during the 35 thousand years of Siberia's population history (Sikora et al., 2019).

Traces of later contacts between the Neo-Eskimos and Paleo-Asiatic peoples are very strongly recognized by genetic data. The Neo-Eskimos were formed on the basis of two genetic components – the Paleo-Eskimo and the Paleo-Indian ones (Flegontov et al., 2019; Sikora et al., 2019). At that, the Paleo-Indian component of the Neo-Eskimos is well recognized by mtDNA haplogroups (A2a, A2b) and Y chromosome

haplogroups (Q-M3). Therefore, by the presence of these haplogroups, it is possible to estimate the genetic contribution of the Neo-Eskimos. Based on mtDNA markers, the frequency of haplogroups A2a and A2b is very high in the Asian Eskimos and Chukchi, while among other Paleo-Asiatic peoples, these haplogroups were found only in the Koryaks at frequencies ranging from 2.7 to 9.1 %⁴ (Derenko et al., 2023). On the Y chromosome, the Paleo-Indian contribution, marked by haplogroup Q-M3, in the Chukchi and Kamchatkan Koryaks has been estimated to be 11.0 and 6.1 %, respectively⁵. The

⁴ Starikovskaya E.B. Phylogeography of the mitogenomes of indigenous populations of Siberia. Dr. Biol. Sci. Diss. Novosibirsk, 2016. (in Russian).

⁵ Kharkov V.N. Structure and phylogeography of the gene pool of the indigenous population of Siberia according to Y chromosome markers. Dr. Biol. Sci. Diss. Tomsk, 2012. (in Russian).

Q-M3 haplogroup was not detected in the Koryak population we studied; however, the frequency of this haplogroup in the Evens is 3.3 % (see Table 1). The most probable reason for the appearance of the “American” haplogroup Q-M3 in the Evens of the Magadan region is interethnic contacts, either with the Koryaks or directly with the Eskimos or related tribes, which, according to archaeological, ethnographic and linguistic data, could have lived on the Sea of Okhotsk coast as early as in the beginning of the 2nd millennium AD (Burykin, 2001).

The high level of interethnic admixture in Northeastern Siberia, mentioned in a number of studies (Khakhovskaya, 2003; Balanovska et al., 2020a, b), is associated with the economic development of this region, first by Russian explorers and then, in the Soviet period, by numerous migrants, mainly of Eastern European origin. In the present study, we also found a high frequency of Y chromosome haplogroups characteristic of Eastern Europeans (and Russians, in particular): haplogroups R, I and J (Derenko et al., 2006; Balanovsky et al., 2008). In the Koryaks, their frequency was 16.7 %, and 37.8 % in the Evens (see Table 1). Moreover, in the Evens, the diversity of R-M17-haplotypes significantly exceeds that of the C-M217 haplogroup characteristic of the Evens themselves ($V_p = 0.225$ and 0.1 , respectively). Meanwhile, the results of the study of maternally inherited mtDNA variability in the Koryaks and Evens of the Magadan region showed that they have a very low frequency of European mtDNA variants (up to 4 % in the Evens) (Derenko et al., 2023). The obtained results, thus, testify to a long history of admixture between the indigenous and immigrant populations in the territory of the Magadan region, as well as to the fact that immigrant men were predominantly involved in interethnic marriages and most of the children of such marriages were likely to be registered as indigenous, which is also typical of other areas of Northeastern Siberia according to demographic data (Khakhovskaya, 2003; Balanovska et al., 2020b).

Conclusion

The results of the study have shown that the male gene pools of the indigenous populations of the Magadan region – the Koryaks and Evens – differ significantly in their structure. The Koryaks have a specific set of Y chromosome haplogroups similar to those of the indigenous peoples of Northeastern Siberia: C-B90-B91, N-B202, Q-B143, while the Evens are characterized by a high frequency of haplogroup C-B80, common among the Tungus-Manchurian peoples. The haplogroups common to the Koryaks and Evens (such as R-M17 and I-P37.2) were obtained from Eastern European migrants as a result of interethnic admixture. The high frequency of this kind of Y chromosome haplogroups in the indigenous peoples of the Magadan region testifies to rather intensive interethnic contacts, mainly from the side of Eastern European males. The analysis of the evolutionary age of aboriginal Y chromosome haplogroups has shown that the gene pools of the Koryaks and Evens are represented by relatively young phylogenetic branches. In the Koryaks, the age of the oldest component of the gene pool (haplogroup C-B91) is estimated to be about 3.8 thousand years; later, haplogroups Q-B143 (about 2.8 thousand years ago) and N-B202 (about 2.4 thousand years ago) appeared in the Koryak gene pool. The Q-B143

haplogroup was most likely inherited by the ancestors of the Koryaks (as well as other Paleo-Asiatic peoples) from the Paleo-Eskimos as a result of their migrations along the Sea of Okhotsk coast. The Evens appeared in the Northern Priokhotye much later (in the XVII century) as a result of the expansion of Tungusic-speaking populations, which is confirmed by the results of the analysis of haplogroup C-B80 polymorphism.

References

- Agdzhoyan A.T., Bogunov Y.V., Bogunova A.A., Kamenshchikova E.N., Kagazezheva Zh.A., Korotkova N.A., Chernyshenko D.N., Ponomarev G.Y., Utrivan S.A., Koshel S.M., Balanovsky O.P., Balanovska E.V. The genetic portrait of the Okhotsk and the Kamchatka Evens population. *Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Moscow University Anthropology Bulletin*. 2019;2: 116-125. DOI 10.32521/2074-8132.2019.2.116-125 (in Russian)
- Agdzhoyan A.T., Bogunova A.A., Kamenshchikova E.N., Zaporozhchenko V.V., Bogunov Y.V., Balanovsky O.P., Balanovska E.V. The Chukchi of Kamchatka: a genetic portrait based on the wide array of Y-chromosome markers. *Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Moscow University Anthropology Bulletin*. 2021;1:80-92. DOI 10.32521/2074-8132.2021.1.080-092 (in Russian)
- Balanovska E.V., Bogunov Y.V., Kamenshchikova E.N., Balaganskaya O.A., Agdzhoyan A.T., Bogunova A.A., Skhalyakho R.A., Alborova I.E., Zhabagin M.K., Koshel S.M., Daragan D.M., Borisova E.B., Galakhova A.A., Maltceva O.V., Mustafin Kh.Kh., Yankovsky N.K., Balanovsky O.P. Demographic and genetic portraits of the Ulchi population. *Russian Journal of Genetics*. 2018;54(10): 1245-1253. DOI 10.1134/S1022795418100046
- Balanovska E.V., Bogunov Y.V., Bogunova A.A., Kamenshchikova E.N., Chernishenko D.N., Pylev V.Y., Balanovsky O.P., Lavryashina M.B. Demographic situation in Chukchi settlements from North Kamchatka. *Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Moscow University Anthropology Bulletin*. 2020a; 1:87-97. DOI 10.32521/2074-8132.2020.1.087-097 (in Russian)
- Balanovska E.V., Bogunov Y.V., Bogunova A.A., Kamenshchikova E.N., Pylev V.Y., Bychkovskaya L.S., Balanovsky O.P., Lavryashina M.B. Demographic portrait of Koryaks from Northern Kamchatka. *Vestnik Moskovskogo Universiteta. Seria XXIII. Antropologia = Moscow University Anthropology Bulletin*. 2020b;4:111-122. DOI 10.32521/2074-8132.2020.4.111-122 (in Russian)
- Balanovsky O., Rootsi S., Pshenichnov A., Kivisild T., Churnosov M., Evseeva I., Pocheshkhova E., Boldyreva M., Yankovsky N., Balanovska E., Villems R. Two sources of the Russian patrilineal heritage in their Eurasian context. *Am. J. Hum. Genet.* 2008;82(1):236-250. DOI 10.1016/j.ajhg.2007.09.019
- Ballantyne K.N., Goedbloed M., Fang R., Schaap O., Lao O., Wollstein A., Choi Y., van Duijn K., Vermeulen M., Brauer S., Decorte R., Poetsch M., von Wurmb-Schwark N., de Knijff P., Labuda D., Vézina H., Knoblauch H., Lessig R., Roewer L., Ploski R., Dobosz T., Henke L., Henke J., Furtado M.R., Kayser M. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.* 2010;87(3):341-353. DOI 10.1016/j.ajhg.2010.08.006
- Bosch E., Calafell F., Rosser Z.H., Nørby S., Lynnerup N., Hurler M.E., Jobling M.A. High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum. Genet.* 2003;112(4):353-363. DOI 10.1007/s00439-003-091
- Burykin A.A. Traces of Eskimo culture on the coast of the Sea of Okhotsk according to archaeological, ethnographic, folkloristic and linguistic data. *Sibirskaya Zaimka*. 2001. URL: <http://zaimka.ru/burykin-eskimos/> (Reference date: 01.09.2023). (in Russian)
- Cardona A., Pagani L., Antao T., Lawson D.J., Eichstaedt C.A., Yngvadottir B., Shwe M.T.T., Wee J., Romero I.G., Raj S., Metspalu M.,

- Villems R., Willerslev E., Tyler-Smith C., Malyarchuk B.A., Derenko M.V., Kivisild T. Genome-wide analysis of cold adaption in indigenous Siberian populations. *PLoS One*. 2014;9(5):e98076. DOI 10.1371/journal.pone.0098076
- Derenko M.V., Malyarchuk B.A. Molecular Phylogeography of Northern Eurasia Populations Based on Mitochondrial DNA Variability Data. Magadan: Far Eastern Scientific Center, Far Eastern Branch of the RAS, 2010 (in Russian)
- Derenko M., Malyarchuk B., Denisova G., Wozniak M., Dambueva I., Dorzhu C., Luzina F., Miscicka-Sliwka D., Zakharov I. Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altay-Sayan regions. *Hum. Genet.* 2006;118(5):591-604. DOI 10.1007/s00439-005-0076-y
- Derenko M., Denisova G., Litvinov A., Dambueva I., Malyarchuk B. Mitogenomics of the Koryaks and Evens of the northern coast of the Sea of Okhotsk. *J. Hum. Genet.* 2023;68(10):705-712. DOI 10.1038/s10038-023-01173-x
- Dulik M.C., Owings A.C., Gaieski J.B., Vilar M.G., Andre A., Lennie C., Mackenzie M.A., Kritsch I., Snowshoe S., Wright R., Martin J., Gibson N., Andrews T.D., Schurr T.G., Genographic Consortium. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc. Natl. Acad. Sci. USA*. 2012;109(22):8471-8476. DOI 10.1073/pnas.1118760109
- Fedorova S.A., Reidla M., Metspalu E., Metspalu M., Rootsi S., Tambets K., Trofimova N., Zhadanov S.I., Hooshiar Kashani B., Olivieri A., Voevoda M.I., Osipova L.P., Platonov F.A., Tomsy M.I., Khusnutdinova E.K., Torroni A., Villems R. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* 2013;13:127. DOI 10.1186/1471-2148-13-127
- Flegontov P., Altınışık N.E., Changmai P., Rohland N., Mallick S., Adamski N., Bolnick D.A., Broomandkhoshbacht N., Candilio F., Culleton B.J., Flegontova O., Friesen T.M., Jeong C., Harper T.K., Keating D., Kennett D.J., Kim A.M., Lamnidis T.C., Lawson A.M., Olalde I., Oppenheimer J., Potter B.A., Raff J., Sattler R.A., Skoglund P., Stewardson K., Vajda E.J., Vasilyev S., Veselovskaya E., Hayes M.G., O'Rourke D.H., Krause J., Pinhasi R., Reich D., Schiffels S. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature*. 2019;570(7760):236-240. DOI 10.1038/s41586-019-1251-y
- Gorin I., Balanovsky O., Kozlov O., Koshel S., Kostryukova E., Zhabagin M., Agdzhoyan A., Pylev V., Balanovska E. Determining the area of ancestral origin for individuals from North Eurasia based on 5,229 SNP markers. *Front. Genet.* 2022;13:902309. DOI 10.3389/fgene.2022.902309
- Grebenyuk P.S., Fedorchenko A.Y., Lebedintsev A.I., Malyarchuk B.A. The ancient cultures of the extreme Northeast Asia and ethnogenetic reconstructions. *Tomskiy Zhurnal Lingvisticheskikh i Antropologicheskikh Issledovaniy = Tomsk Journal of Linguistics and Anthropology*. 2019;2(24):110-136. DOI 10.23951/2307-6119-2019-2-110-136 (in Russian)
- Grugni V., Raveane A., Ongaro L., Battaglia V., Trombetta B., Colombo G., Capodiferro M.R., Olivieri A., Achilli A., Perego U.A., Motta J., Tribaldoni M., Woodward S.R., Ferretti L., Cruciani F., Torroni A., Semino O. Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. *BMC Biol.* 2019;17(1):3. DOI 10.1186/s12915-018-0622-4
- Ilumäe A.M., Reidla M., Chukhryaeva M., Järve M., Post H., Karmin M., Saag L., Agdzhoyan A., Kushniarevich A., Litvinov S., Ekomasova N., Tambets K., Metspalu E., Khusainova R., Yunusbayev B., Khusnutdinova E.K., Osipova L.P., Fedorova S., Utevska O., Koshel S., Balanovska E., Behar D.M., Balanovsky O., Kivisild T., Underhill P.A., Villems R., Rootsi S. Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families. *Am. J. Hum. Genet.* 2016;99(1):163-173. DOI 10.1016/j.ajhg.2016.05.025
- Karafet T.M., Osipova L.P., Savina O.V., Hallmark B., Hammer M.F. Siberian genetic diversity reveals complex origins of the Samoyedic-speaking populations. *Am. J. Hum. Biol.* 2018;30(6):e23194. DOI 10.1002/ajhb.23194
- Karmin M., Saag L., Järve M., Vicente M., Wilson Sayres M.A., Pagan L., DeGiorgio M., Talas U.G., Rootsi S., Ilumäe A.M., ... Tyler-Smith C., Underhill P., Willerslev E., Nielsen R., Metspalu M., Villems R., Kivisild T. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 2015;25(4):459-466. DOI 10.1101/gr.186684.114
- Kayser M., Krawczak M., Excoffier L., Dieltjes P., Corach D., Pascoli V., Gehrig C., Bernini L.F., Jespersen J., Bakker E., Roewer L., de Knijff P. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am. J. Hum. Genet.* 2001;68(4):990-1018. DOI 10.1086/319510
- Khakhovskaya L.N. Kamchadals of Magadan Oblast: history, culture, identification. Magadan: North-Eastern Interdisciplinary Research Institute, Far Eastern Branch of the RAS, 2003 (in Russian)
- Luis J.R., Palencia-Madrid L., Garcia-Bertrand R., Herrera R.J. Bidirectional dispersals during the peopling of the North American Arctic. *Sci. Rep.* 2023;13(1):1268. DOI 10.1038/s41598-023-28384-8
- Malyarchuk B., Derenko M., Denisova G., Maksimov A., Wozniak M., Grzybowski T., Dambueva I., Zakharov I. Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. *J. Hum. Genet.* 2011;56(8):583-588. DOI 10.1038/jhg.2011.64
- Malyarchuk B., Derenko M., Denisova G., Khoit S., Wozniak M., Grzybowski T., Zakharov I. Y-chromosome diversity in the Kalmyks at the ethnical and tribal levels. *J. Hum. Genet.* 2013;58(12):804-811. DOI 10.1038/jhg.2013.108
- Olofsson J.K., Pereira V., Børsting C., Morling N. Peopling of the north circumpolar region – insights from Y chromosome STR and SNP typing of Greenlanders. *PLoS One*. 2015;10(1):e0116573. DOI 10.1371/journal.pone.0116573
- Pagan L., Lawson D.J., Jagoda E., Mörseburg A., Eriksson A., Mitt M., Clemente F., Hudjashov G., DeGiorgio M., Saag L., ... Thomas M.G., Manica A., Nielsen R., Villems R., Willerslev E., Kivisild T., Metspalu M. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature*. 2016;538(7624):238-242. DOI 10.1038/nature19792
- Pakendorf B., Novgorodov I.N., Osakovskij V.L., Danilova A.P., Protop'jakonov A.P., Stoneking M. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum. Genet.* 2006;120(3):334-353. DOI 10.1007/s00439-006-0213-2
- Pugach I., Matveev R., Spitsyn V., Makarov S., Novgorodov I., Osakovskiy V., Stoneking M., Pakendorf B. The complex admixture history and recent southern origins of Siberian populations. *Mol. Biol. Evol.* 2016;33(7):1777-1795. DOI 10.1093/molbev/msw055
- Raghavan M., Skoglund P., Graf K.E., Metspalu M., Albrechtsen A., Moltke I., Rasmussen S., Stafford T.W., Jr., Orlando L., Metspalu E., Karmin M., Tambets K., Rootsi S., Mägi R., Campos P.F., Balanovska E., Balanovsky O., Khusnutdinova E., Litvinov S., Osipova L.P., Fedorova S.A., Voevoda M.I., DeGiorgio M., Sichevitz-Ponten T., Brunak S., Demeshchenko S., Kivisild T., Villems R., Nielsen R., Jakobsson M., Willerslev E. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. 2014;505(7481):87-91. DOI 10.1038/nature12736
- Rasmussen M., Li Y., Lindgreen S., Pedersen J.S., Albrechtsen A., Moltke I., Metspalu M., Metspalu E., Kivisild T., Gupta R., ... Brunak S., Sichevitz-Pontén T., Villems R., Nielsen R., Krogh A., Wang J., Willerslev E. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. 2010;463(7282):757-762. DOI 10.1038/nature08835

- Rubicz R., Zlojutro M., Sun G., Spitsyn V., Deka R., Young K.L., Crawford M.H. Genetic architecture of a small, recently aggregated Aleut population: Bering Island, Russia. *Hum. Biol.* 2010;82(5-6): 719-736. DOI 10.1353/hub.2010.a408138
- Sikora M., Pitulko V., Sousa V., Allentoft M.E., Vinner L., Rasmussen S., Margaryan A., de Barros Damgaard P., de la Fuente Castro C., Renaud G., ... Sajantila A., Lahr M.M., Durbin R., Nielsen R., Meltzer D., Excoffier L., Willerslev E. The population history of north-eastern Siberia since the Pleistocene. *Nature.* 2019;570(7760):182-188. DOI 10.1038/s41586-019-1279-z
- Solovyev A.V., Borisova T.V., Romanov G.P., Teryutin F.M., Pshenikova V.G., Nikitina S.E., Alekseev A.N., Barashkov N.A., Fedorova S.A. Genetic history of Russian Old-Settlers of the Arctic coast of Yakutia from the settlement of Russkoye Ust'ye inferred from Y chromosome data and genome-wide analysis. *Russian Journal of Genetics.* 2023;59(9):949-955. DOI 10.1134/S1022795423090119
- Yu H.-X., Ao C., Zhang X.-P., Liu K.-J., Wang Y.-B., Meng S.-L., Li H., Wei L.-H., Man D. Unveiling 2,000 years of differentiation among Tungusic-speaking populations: a revised phylogeny of the paternal founder lineage C2a-M48-SK1061. *Front. Genet.* 2023;14:1243730. DOI 10.3389/fgene.2023.1243730
- Zhivotovsky L.A., Underhill P.A., Cinnioğlu C., Kayser M., Morar B., Kivisild T., Scozzari R., Cruciani F., Destro-Bisol G., Spedini G., Chambers G.K., Herrera R.J., Yong K.K., Gresham D., Tournev I., Feldman M.W., Kalaydjieva L. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 2004;74(1):50-61. DOI 10.1086/380911

ORCID

B.A. Malyarchuk orcid.org/0000-0002-0304-0652
M.V. Derenko orcid.org/0000-0002-1849-784X

Acknowledgements. The study was supported by a grant from the Russian Science Foundation No. 22-24-00264 (<https://rscf.ru/project/22-24-00264/>).

Conflict of interest. The authors declare no conflict of interest.

Received September 14, 2023. Revised November 29, 2023. Accepted November 29, 2023.

Original Russian text <https://vavilovj-icg.ru/>

A universal panel of STR loci for the study of polymorphism of the species *Canis lupus* and forensic identification of dog and wolf

A.E. Hrebianchuk¹✉, I.S. Tsybovsky²

¹ Forensic Examination Committee of the Republic of Belarus, Minsk, Belarus

² BelJurZabespjachjenne, Minsk, Belarus

✉ iamsanya94@mail.ru

Abstract. Commercial panels of microsatellite (STR) loci are intended for DNA analysis of the domestic dog (*Canis lupus familiaris*) and, therefore, when genotyping the Grey wolf (*Canis lupus lupus*), most markers reveal significant deviations from the Hardy–Weinberg equilibrium and have a low informative value, which complicates their use in a forensic examination. The aim of this study was to select STR markers that equally effectively reflect population polymorphism in the wolf and the dog, and to create a universal panel for the identification of individuals in forensic science. Based on the study of polymorphisms of 34 STR loci, a Cplex panel of 15 autosomal loci and two sex loci was developed, which is equally suitable for identifying wolves and dogs. Analysis of molecular variance (AMOVA) between samples revealed significant differentiation values ($F_{ST} = 0.0828$, $p < 0.05$), which allows the panel to be used for differentiating between wolf and dog samples. For the first time in the forensic examination of objects of animal origin in the Republic of Belarus, population subdivision coefficients (θ -values) were calculated for each of the 15 STR loci of the test system being reported. It was shown that the values of the genotype frequency, when averaged over all studied animals without and with considering the θ -value, differ by three orders of magnitude ($3.39 \cdot 10^{-17}$ and $4.71 \cdot 10^{-14}$, respectively). The use of population subdivision coefficients will provide the researcher with the most relevant results of an expert identification study. The test system was validated in accordance with the protocol of the Scientific Working Group on DNA Analysis Methods. A computational tool was developed to automate the analysis of genetic data on the wolf and dog in the forensic examination; two guides were approved for practicing forensic experts. This methodology is being successfully used in expert practice in investigating cases of illegal hunting, animal abuse and other offenses in the Republic of Belarus.

Key words: microsatellites; polymorphism; differentiation; identification; *Canis lupus familiaris*; *Canis lupus lupus*; wildlife forensic science.

For citation: Hrebianchuk A.E., Tsybovsky I.S. A universal panel of STR loci for the study of polymorphism of the species *Canis lupus* and forensic identification of dog and wolf. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2024;28(1):98-107. DOI 10.18699/vjgb-24-12

Универсальная панель STR-локусов для исследования полиморфизма вида *Canis lupus* и криминалистической идентификации собаки и волка

A.E. Гребенчук¹✉, И.С. Цыбовский²

¹ Государственный комитет судебных экспертиз Республики Беларусь, Минск, Беларусь

² Республиканское унитарное предприятие «БелЮрОбеспечение», Минск, Беларусь

✉ iamsanya94@mail.ru

Аннотация. Коммерческие панели микросателлитных (STR) локусов предназначены для работы с ДНК собаки домашней (*Canis lupus familiaris*), в связи с чем при генотипировании волка обыкновенного (*Canis lupus lupus*) большинство маркеров показывают существенные отклонения от равновесия Харди–Вайнберга и имеют низкий показатель информативной ценности, что осложняет их использование в судебной экспертизе. Целью настоящего исследования стал подбор STR-маркеров, которые одинаково эффективно отображают популяционный полиморфизм волка и собаки, с последующим созданием универсальной панели для дифференциации и идентификации особей волка и собаки в криминалистике. На основе исследования полиморфизма 34 STR-локусов сконструирована панель Cplex из 15 аутомомных локусов и двух локусов половой принадлежности, которая одинаково применима для идентификации волка и собаки. Анализ молекулярной дисперсии (AMOVA) между выборками выявил достоверные значения дифференциации ($F_{ST} = 0.0828$, $p < 0.05$), что позволяет использовать панель для дифференциации образцов волка обыкновенного и собаки домашней. Впервые в судебной экспертизе объектов животного происхождения в Республике Беларусь рассчитаны коэффициенты подразделенности популяции (θ -value) для каждого из 15 STR-локусов разработанной тест-системы. Показано, что значения частот генотипов, усреднен-

ные по всем исследованным животным без учета и с учетом θ -value, различаются на три порядка ($3.39 \cdot 10^{-17}$ и $4.71 \cdot 10^{-14}$ соответственно). Применение коэффициентов подразделенности популяции позволит оперировать наиболее достоверными результатами экспертного идентификационного исследования. Предложенная тест-система валидирована в соответствии с протоколом Scientific Working Group on DNA Analysis Methods. Создан информационно-статистический комплекс для автоматизации обчета генетических данных волка обыкновенного и собаки домашней в судебной экспертизе, утверждены две методики для практикующих судебных экспертов. Методические разработки успешно применяются в экспертной практике при расследовании фактов незаконной охоты, жестокого обращения с животными и других правонарушений в Республике Беларусь.

Ключевые слова: микросателлиты; полиморфизм; дифференциация; идентификация; собака домашняя; волк обыкновенный; судебная экспертиза объектов животного происхождения.

Introduction

According to the statistics of the Ministry of Forestry, in the Republic of Belarus, the Grey wolf (*Canis lupus lupus*) population has stabilized over the past five years in the range of 1,530–1,630 individuals, which is one of the leading indicators among European countries (Ministry of Forestry of the Republic of Belarus, 2021). Notably, wolf hunting in Belarus is allowed all year long with no sex or age restrictions. At the same time, hunting in protected natural areas and hunting without a permit results in criminal cases and, consequently, the need for forensic examination.

According to the Ministry of Housing and Communal Services, about 80,000 stray cats and dogs are exterminated in Belarus every year, and this number is growing, while the exact number of dogs is unknown. Being one of the most common companion animals, the domestic dog (*Canis lupus familiaris*) has a special status among farm and domestic animals. The active use of dogs by humans in various roles is also reflected in the criminal aspects that accompany the development of society.

The natural history of European wolf (*C. lupus lupus*) populations has been characterized by a strong reduction in the number of individuals over the past few hundred years (Boitani, 2003). The decline of the population, its fragmentation, and disruption of the gene flow are well-known triggers of genetic impoverishment and increased inbreeding in natural populations, which increase the risk of extinction for wolves as well as for many other species. An example of such a situation was documented for wolves in Italy, where the values of genetic diversity determined by the level of heterozygosity were clearly lower than those in populations from Russia, Alaska, and Canada (Godinho et al., 2011).

Intensive extermination of the wolf can lead to the replacement of the wolf with wolf-dog hybrids. Recently, the problem of hybridization between wolves and free-living dogs in Europe has become a major topic in many research programs (Stronen et al., 2022).

The main difficulty in the genetic differentiation between the wolf and the dog is that DNA markers unique to both the wolf and the dog have not been found. A comparison of dog and wolf genomes showed a similarity of 99 %, which once again confirms their common origin (Freedman, Wayne, 2017).

The study of wolf populations is usually designed according to a typical approach that includes the use of loci recommended by the International Society of Animal Genetics (ISAG) with calculation of statistical indexes of distribution of alleles of the studied loci and assessment of the representation of alleles in the population. Due to the high level of identity of wolf and dog genomes, the ranges of alleles of the loci are very similar.

Therefore, differentiation of individuals using selected loci becomes possible if genetic differentiation of the studied samples is revealed by statistical processing of genotyping results (Halverson, Basten, 2005; Fan et al., 2016).

Most panels for canine DNA analysis are unsuitable for the study of wolf DNA due to deviation from the Hardy–Weinberg equilibrium and the presence of null alleles in DNA markers. On the other hand, when selecting markers for DNA analysis of the wolf, researchers generally do not take into account the possibility of using the selected markers on an inbred dog population, which precludes the use of these markers to identify a wolf and a dog simultaneously in forensic studies.

The aim of this study was to select STR markers that equally effectively reflect the population structure and the polymorphism of the Grey wolf and the domestic dog, followed by the creation of a universal panel for the identification and differentiation of individuals in forensic science.

Materials and methods

Biological objects. Biological samples of the Grey wolf and the domestic dog were obtained legally and are represented by fragments of muscle and cartilage tissues of wolves ($n = 103$) and samples of blood, hair and saliva of purebred dogs, mestizo dogs and outbred dogs ($n = 198$). The list of the most represented breeds is reflected in Supplementary Material 1¹.

DNA extraction, amplification and genotyping of microsatellite loci. DNA from muscle and cartilage tissues, blood, saliva and hair of wolves and dogs was extracted according to a method based on the release of DNA during the incubation of biological material in a lysis buffer with proteinase K and 0.01 mM dithiothreitol at 37–56 °C. The lysate was subjected to the purification procedure on silica (Boom et al., 1990).

All selected loci were grouped into two test systems: a test system that includes mainly dinucleotide loci recommended by ISAG: AHTk211, FH2054, CXX279, Ren169O18, INU055, AHTTh260, INU030, FH2079, FH2848, AHT121, AHTTh171, Ren247M23, AHTTh130, INRA21, AHTk253, AHT137, Ren54P11, INU005, Ren105L03, Ren64E19, Ren162C04 and Amelogenin sex locus (Radco, Podbielska, 2021); and a test system that includes mainly tetranucleotide loci – FH2096, CPH12, CPH4, FH2004 (Caniglia et al., 2010), FH2016 (Fan et al., 2016), FH2361, PEZ17, FH2328 (van Asch et al., 2010), PEZ16, vWF.x (DeNise et al., 2004), FH2010 (Eichmann et al., 2004), FH2001 (Verardi et al., 2006), VGL3438 (Magory Cohen et al., 2013) and sex loci – DBX and DBY (Seddon, 2005).

¹ Supplementary Materials 1–6 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Hrebian_Engl_28_1.pdf

Accordingly, the 10 µl PCR reaction volume contained 10 mM tris-HCl, pH 8.6; 25 mM KCl; 2.0 mM MgCl₂; 0.2 mM of each dNTP; 0.2–1.0 µM of each of the pair of primers; 0.15 u. DNA polymerase; 1.5 ng/µl BSA; 0.1 % Triton X-100 and 1–20 ng DNA to be analyzed. The polymerase chain reaction was conducted in a thermocycler C1000 (BioRad, USA) using the following program: (1) an initial denaturation step at 95 °C for 5 min; (2) 30 cycles of denaturation at 95 °C (for 30 s), annealing at 60 °C (for 40 s) and elongation at 72 °C (for 1 min); (3) a final elongation at 72 °C for 30 min.

The combination of alleles of each of the samples was detected by electrophoretic separation of PCR products on a 3500 Genetic Analyzer (ThermoFisher Scientific, USA). The size of the detected alleles (in bp) in the studied loci was determined using the Orange 500 bp internal size standards (NimaGen®, The Netherlands) and GeneScan-600 LIZ™ SizeStandard v2.0 (ThermoFisher Scientific, USA). Genotyping was evaluated with GeneMapper ID-X v1.6 software package (ThermoFisher Scientific, USA).

Statistical analysis of the data. Since incorrect species identification distorts calculations based on the analysis of genetic diversity indexes (Galinskayaa et al., 2019), a cluster analysis of genotyping data for wolves and dogs was first carried out for the studied loci. The population structure was determined using the Monte Carlo algorithm according to the Markov chain method using STRUCTURE v.2.3.4 software and Admixture model (Pritchard et al., 2000) with further determination of the true number of clusters by the method of Evanno et al. (Evanno et al., 2005). The burn-in period included 500,000 iterations, followed by the construction of the Markov chain for 1,000,000 iterations for the expected number of groups in the sample, K , from 1 to 10, with six repeats for each value of K . The analysis of clustering in the combined pool of wolves and dogs was performed using the unweighted pair group method with arithmetic mean (UPGMA) and the nearest neighbors joining method (NJ), and the construction of the corresponding dendrograms was carried out using MEGA v.11.0.10 software. To visualize the genetic structure, a multidimensional analysis was performed using the genetic distance matrix following the PCoA method in GenAlEx v.6.5 (Peakall, Smouse, 2006, 2012).

The calculation of frequencies of alleles of STR loci and values of the observed (H_O) and expected heterozygosity (H_E), as well as evaluation of deviation from Hardy–Weinberg equilibrium was performed using Cervus v.3.0.7 software package (Kalinowski et al., 2007). To identify possible errors in the interpretation of genetic profiles, null alleles, and PCR artifacts, an analysis was carried out using Micro-Checker v.2.2.1 software (van Oosterhout et al., 2004).

Analysis of molecular dispersion and estimation of inbreeding coefficients were performed using Arlequin v.3.5.1.3 software (Excoffier et al., 2005). Analysis of assignment of the individual to the sample pool (Assignment Test) based on the selected loci was performed using GenAlEx v.6.5. Polymorphism information content (PIC) of the selected STR loci was calculated using Cervus software v.3.0.7.

Alleles sequence determination. To identify possible microvariants of the sequence, as well as to perform tandem determination of alleles, which is a common practice in

forensic science, the primary structure of alleles was determined by Sanger dideoxy sequencing (Sanger et al., 1977). Nucleotide sequences of alleles of each STR locus and sex loci were determined in the forward and reverse directions. Sequencing was performed on a 3500 Genetic Analyzer and was conducted with the BigDye® Terminator v.3.1 Cycle Sequencing Kit (ThermoFisher Scientific). Comparative analysis of allele sequences of the studied loci was carried out in BioEdit v.7.0.5.3 (Hall et al., 2011). The sequences of each locus with the minimum and maximum allele molecular sizes were deposited in the GenBank database (Benson et al., 2005) with assignment of corresponding access numbers.

Results and discussion

A posteriori analysis of the STRUCTURE results for the combined pool of the wolf and dog genotypes revealed the maximum value of the test statistics ΔK at $K = 2$, which indicates the presence of two genetic clusters in the analyzed sample of animals: the Grey wolf (green cluster) and the domestic dog (red cluster) (Fig. 1, Supplementary Material 2). When determining the population structure of samples of wolves and dogs separately from each other, the cluster formed by the wolf samples remained homogeneous; the absence of clustering within the wolf population using STR loci was previously shown by researchers in Europe (Aspi et al., 2006; Sastre et al., 2011; Ðan et al., 2016).

For a separate analysis of STRUCTURE in the samples of dogs, four groups were formed: purebred dogs ($n = 78$) and three breed groups: Molossians ($n = 32$), Samoyeds ($n = 66$) and Shepherd dogs ($n = 22$). The maximum value of ΔK at $K = 2$ indicated the formation of two clusters (Supplementary Materials 3 and 4), one of which corresponds to Shepherd dogs, whereas the breed groups of Samoyeds and Molossians were not separated and, moreover, did not differ from the group of outbred dogs.

At the same time, the analysis of clustering in the combined pool of wolves and dogs by the methods of UPGMA and NJ with construction of corresponding dendrograms revealed four clusters of different hierarchical levels in the sample of dogs. The greatest similarity was observed between the breed group of Molossians and outbred dogs. They are adjacent to the breed group of Samoyeds, and this entire group is separated from the group of Shepherd dogs. Since both dendrograms showed a similar structure, Supplementary Material 5 shows only the dendrogram built according to the UPGMA method (with bootstrapping for 10,000 permutations). It should be noted that a similar pattern of clustering of dog breeds in the study of SNP markers was obtained by other authors (Parker et al., 2017).

Analysis of the population structure of wolves and dogs showed a strong genetic differentiation between them with the average values of the cluster membership coefficient Q of 0.984 and 0.981, respectively. The data obtained on differentiation of the wolf and the dog by 34 STRs are in good agreement with the data of Korablev et al. (2021). The level of differentiation between dog breeds was much lower than between the wolf and the dog. Q values varied from 0.457 to 0.495 for Molossians, from 0.451 to 0.476 for Samoyeds, and from 0.740 to 0.757 for Shepherd dogs.

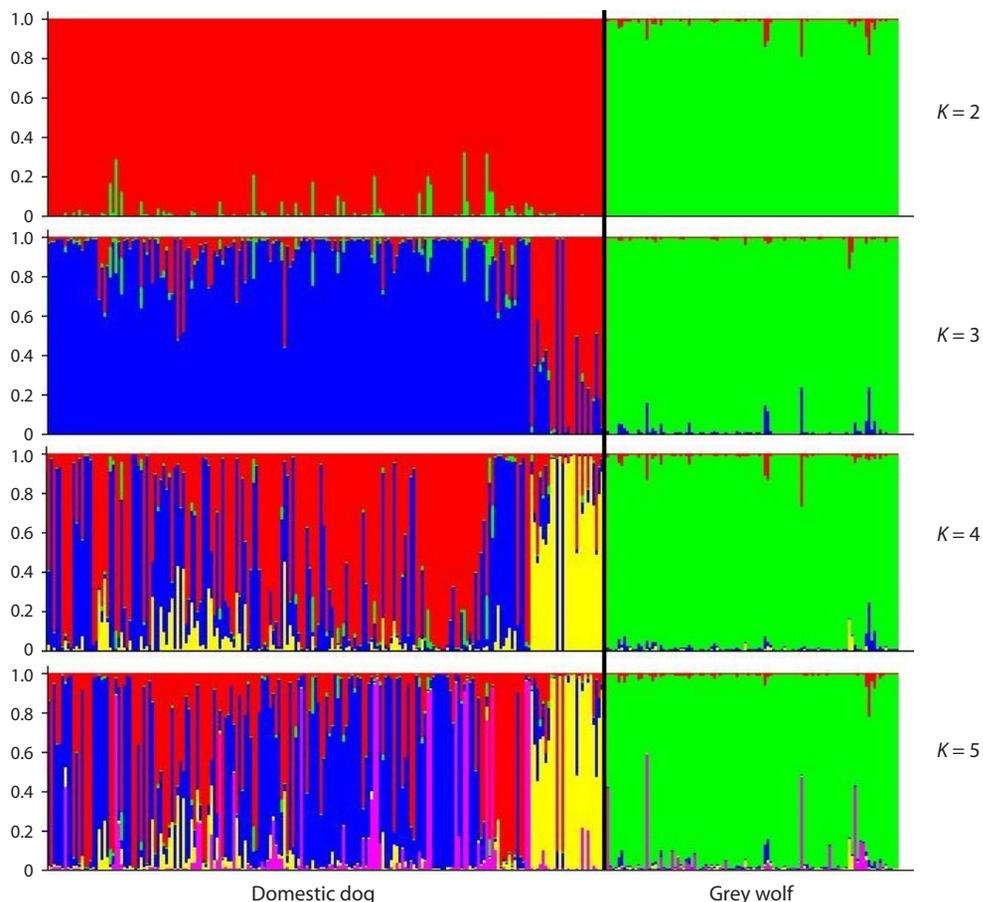


Fig. 1. Results of cluster analysis of wolf and dog samples, performed in STRUCTURE software, for the most probable value of the number of genetic clusters $K = 2-5$, sorted by samples.

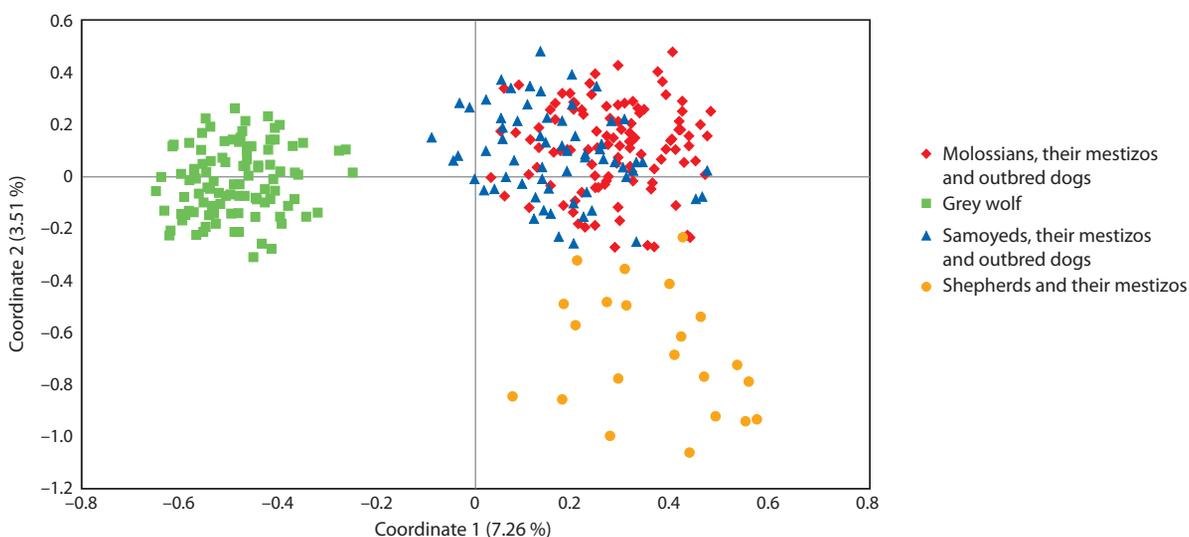


Fig. 2. The diagram of the results of PCoA analysis based on the matrix of paired F_{ST} values for samples of the Grey wolf and domestic dog.

The results of the cluster analysis are consistent with the multivariate genetic distance matrix analysis (PCoA method), which also shows a strong differentiation between the wolf and dog samples (Fig. 2). Similarly, Shepherd dogs form a separate group. The latter outcome may be explained by selection dif-

ferences or may be a consequence of the high heterogeneity of the samples of Molossians and Samoyeds, which have a common historic origin (over 20 different breeds), while the sample of Shepherd dogs included only purebred German Shepherds.

Table 1. Average values of the level of polymorphism of the studied loci in samples of wolves and three historical breed groups of dogs

Samples	N_a	N_e	H_O	H_E
Wolves	9.265 ± 0.476	5.258 ± 0.306	0.730 ± 0.017	0.786 ± 0.013
Molossians, their mestizos and outbred dogs	9.706 ± 0.575	4.916 ± 0.293	0.656 ± 0.017	0.768 ± 0.016
Samoyeds, their mestizos and outbred dogs	9.235 ± 0.560	5.212 ± 0.312	0.688 ± 0.018	0.782 ± 0.015
Shepherds and their mestizos	5.824 ± 0.328	2.992 ± 0.197	0.556 ± 0.029	0.617 ± 0.025

Note. N_a – number of alleles per locus; N_e – effective number of alleles per locus; H_O – observed heterozygosity; H_E – expected heterozygosity.

In total, 405 alleles were identified in the sample of wolves and domestic dogs using the two test systems. All the loci were polymorphic and had from 5 (FH2096) to 26 (FH2361) alleles per locus. The average number of alleles per locus in all samples was similar, amounting to 9.402 ± 0.617 (Table 1). The Shepherd dog was an exception, with the mean value per locus in the range of 5.824 ± 0.328 , which could be a consequence of the small sample size ($n = 22$).

The highest observed and expected heterozygosity rates were obtained for the wolf sample, at 0.730 and 0.786, respectively (see Table 1). In dog samples, lower values of expected heterozygosity compared to observed heterozygosity are a potent indicator of the presence of inbreeding resulting in synthetic selection and genetic drift, which can irreversibly remove alleles from the population, leading to significantly reduced diversity (Galinskayaa et al., 2019). Although mutations counteract the genetic drift, it is difficult to achieve a balance of genetic processes in dog breeding due to exclusion of individuals with identified mutations in a particular trait. The highest values of heterozygosity and the highest effective number of alleles were found in the wolf sample, which indicates natural development and presence of mutation-drift balance in the natural population.

The analysis of the profiles of dinucleotide loci requires special attention since fragments that do not belong to true alleles and are stutter products can be present on electrophoregrams. The percentage of stutter products usually increases with the length of the allele. Based on the analysis of the genotypes using Micro-Checker software, three of the 20 dinucleotide markers (INU055, Ren169O18, and Ren64E19) showed a high probability of genotyping errors and were excluded from further analysis.

The analysis of the distribution of alleles for two loci (AHT121 and AHTk211) showed the presence of a large number of null alleles in three samples, including the sample of the wolf. AHT137 and INRA21 loci showed a high null alleles content in two samples of dogs, 8 loci (AHT130, AHT260, CXX279, FH2848, Ren105L03, Ren162C04, Ren247M23, and Ren54P11) showed a rather high null alleles content (from 5.7 % in the sample of Molossians at the FH2848 locus to 12.6 % at the CXX279 locus in the sample of Shepherd dogs). These loci were also excluded from further work on the design of a universal forensic panel.

When analyzing the pattern of frequency distribution of alleles in specific loci, special attention was paid to loci with a significant predominance of major alleles. Differences in the frequencies of major alleles can be very instrumental for

differentiating wolves and dogs by microsatellite analysis; however, a significant predominance of one allele can affect the level of identification confidence. Due to the pronounced dominance of major alleles, INU030, INU005, AHTk171 and AHTk253 loci were excluded from further analysis.

Most of the studied loci in the samples of wolves and dogs conformed to the Hardy–Weinberg distribution ($p > 0.05$), including AHT171, AHTk253, CPH12, CPH4, FH2001, FH2004, FH2010, FH2016, FH2096, FH2328, FH2361, INU005, INU030, INU055, PEZ16, PEZ17, Ren169O18, Ren64E19, VGL3438 and vWF.x. Two loci, FH2054 and FH2079, deviated from the equilibrium ($p = 0.005$) in all studied samples; however, when using the Bonferroni correction, p -values ceased to be statistically significant. In 12 loci, a statistically significant deviation from the Hardy–Weinberg equilibrium was revealed in at least one sample, which may reflect the manifestation of loci in the studied samples and can be explained by the presence of null alleles.

The allele fixation index (F_{ST}) ranged from 0.025 (FH2001) to 0.158 (CPH4), with an average value of 0.077 for all loci. The highest F_{IS} values were obtained for the FH2079, FH2096, and CPH12 loci (0.221, 0.174, and 0.162, respectively). Overall, for a panel of 15 selected loci, inbreeding indexes (F_{IS} and F_{IT}) were 0.103 and 0.172, respectively (Table 2).

F_{IS} values in the sample of wolves in most of the studied loci showed values approaching zero, which, together with high heterozygosity values, can indicate the presence of panmixia in the population (Galinskayaa et al., 2019).

A correct assignment of a sample to a Grey wolf or a domestic dog can be of crucial importance in a forensic investigation. Analysis of molecular variance (AMOVA) was performed to assess the possibility of differentiation between a wolf and a dog using selected microsatellite loci. The AMOVA results showed that the percentage of variation between wolf and dog samples was 8.28 %, while within samples it was 91.72 %. Variance components in the population were significant for all studied loci (Supplementary Material 6), which indicates differentiation of the wolf and dog samples. The vWF.x, FH2096, and CPH4 loci accounted for 21.11, 19.34, and 14.40 % of inter-sample genetic variability, respectively, while FH2079 and FH2016 showed the lowest inter-population variability (2.45 and 2.20 %, respectively).

According to Wright's interpretation (Wright, 1978), the range of F_{ST} values from 0.15 to 0.25 indicates moderate differentiation. At the same time, values in the range of 0.00–0.05 indicate a weak but noteworthy difference between the samples. Since hypervariable markers with a large number of

Table 2. Mean values of heterozygosity for the wolf and dog and Wright F-statistic values for the total sample of *Canis lupus*

Loci	H_O	H_E	H_O	H_E	F_{IS}	F_{ST}	F_{IT}
	Domestic dog		Grey wolf				
FH2016	0.783	0.894	0.832	0.886	0.129	0.033	0.158
FH2079	0.555	0.664	0.612	0.656	0.221	0.093	0.293
VGL3438	0.751	0.858	0.709	0.784	0.130	0.029	0.155
FH2361	0.793	0.852	0.861	0.821	0.003	0.032	0.035
FH2054	0.792	0.822	0.760	0.781	0.098	0.077	0.167
FH2001	0.737	0.804	0.612	0.744	0.054	0.025	0.078
FH2328	0.763	0.865	0.718	0.743	0.065	0.058	0.119
FH2004	0.767	0.803	0.784	0.885	0.081	0.067	0.142
CPH12	0.571	0.707	0.553	0.650	0.162	0.060	0.212
FH2010	0.641	0.713	0.767	0.777	0.022	0.070	0.090
PEZ17	0.675	0.797	0.777	0.776	0.109	0.068	0.170
PEZ16	0.611	0.793	0.786	0.825	0.145	0.120	0.247
CPH4	0.595	0.611	0.767	0.837	0.103	0.158	0.245
FH2096	0.414	0.524	0.485	0.603	0.174	0.133	0.284
vWF.X	0.545	0.613	0.786	0.838	0.052	0.136	0.182
Average	0.666	0.755	0.721	0.774	0.103	0.077	0.172

Note. H_O – observed heterozygosity; H_E – expected heterozygosity; F_{IS} – inbreeding index of individuals within the sample; F_{ST} – allele fixation index; F_{IT} – inbreeding index of individuals in the total sample.

alleles can have significantly lower F_{ST} values than markers with a small number of alleles, it is more important to detect significant genetic differentiation between wolves and dogs in the totality of the selected STR loci (Hedrick, 2000). Between-sample AMOVA analysis revealed significant differentiation between the wolf and the dog ($F_{ST} = 0.0828, p < 0.05$).

The significance of the differentiation of the wolf and the dog using the selected loci can be illustrated by the analysis of assignment of a definite sample (Assignment test). The analysis is based on the calculation of the probability value of the presence of the genotype of a certain individual in the sample from which it was selected, and its comparison with the probability value of the same genotype in another sample. Based on these calculations, an individual belongs to the sample for which it has the highest probability (Fig. 3).

The calculation of true genetic affiliation to the sample showed a high consolidation of wolves and dogs, with 100 % of all studied animals genetically assigned to their own cluster. At the same time, we observed a large difference in the moduli of natural logarithms of expected frequencies of the genotypes when they belonged to their own versus an alternative sample; for wolves, the difference was on average 8.722, and for the domestic dog, it was 8.584. The high values of the difference between the moduli of the logarithms of the expected genotype frequencies confirm successful differentiation of the Grey wolf and the domestic dog using the proposed loci.

An important criterion in forensics is adequate interpretation of the value of the reliability level of an identification study. To

calculate the probability of accidental matching of genotypes, an appropriate reference database and correct application of a conservative calculation procedure are required (Buckleton et al., 2016). The study of genotype pools of samples of wolves and dogs made it possible to calculate the frequencies of occurrence of alleles and corrections for the genetic subdivision of the populations of wolves and dogs living in Belarus.

A complex social organization of the wolf pack, along with targeted dog breeding, lead to formation of structured populations of these animals. Therefore, in order to obtain the most reliable conclusions, using frequency characteristics of the loci of specific subpopulations would be the preferred approach in an identification study, but this is hardly possible in practice in forensic examination of objects of animal origin. An alternative solution is to introduce a subdivision coefficient (θ -value) into the calculation of the frequency of the genotype, which takes into account the presence of structured populations (The Evaluation..., 1996; Buckleton et al., 2006).

As can be seen from Table 3, the values of genotype frequencies averaged over all studied animals and calculated with the inclusion of subsequent studied loci of the test system without and with taking into account the θ -value differ by three orders of magnitude ($3.39 \cdot 10^{-17}$ and $4.71 \cdot 10^{-14}$, respectively). This suggests that not factoring in the θ -value in the identification study will lead to overestimation (underestimation) of the genotype frequency, which in forensic examination will lead to an unintended overestimation of the reliability level of the study.

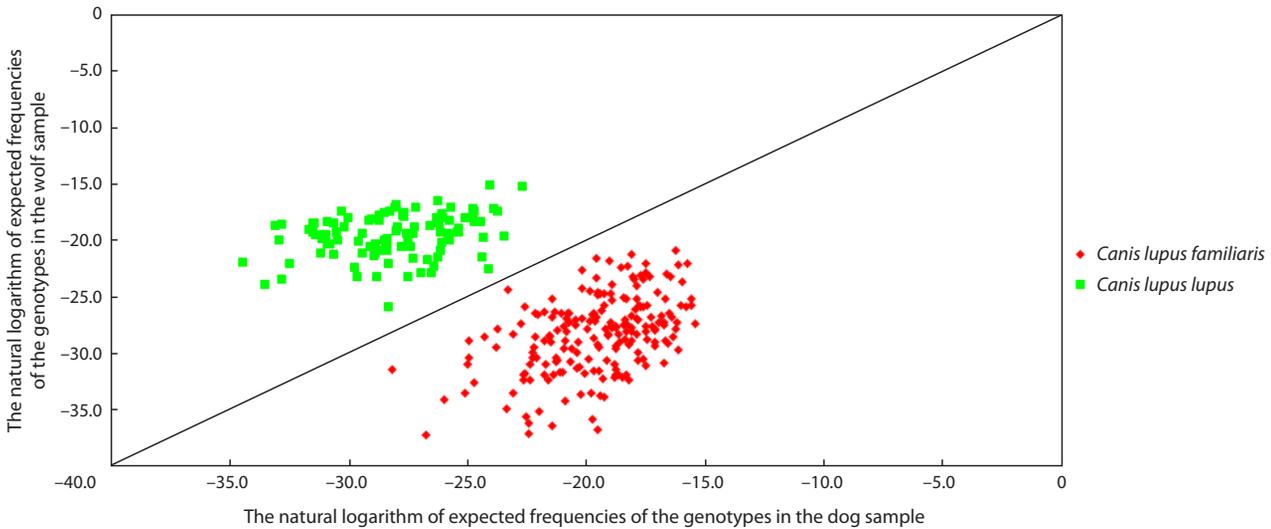


Fig. 3. Graph of the genetic assignment of the genotype to the study sample (graphical interpretation of the Assignment test).

Table 3. Genotype frequencies averaged over all animals, without and with using the θ -value

Loci	θ -value	Genotype frequency excluding θ -value and with the inclusion of the subsequent locus	Genotype frequencies with θ -value and with the inclusion of the subsequent locus**
FH2016	0.039	$2.08 \cdot 10^{-2}$	$3.62 \cdot 10^{-2}$
FH2079	0.145	$3.31 \cdot 10^{-3*}$	$9.58 \cdot 10^{-3**}$
VGL3438	0.020	$1.55 \cdot 10^{-4}$	$5.44 \cdot 10^{-4}$
FH2361	0.035	$7.19 \cdot 10^{-6}$	$3.39 \cdot 10^{-5}$
FH2054	0.081	$4.26 \cdot 10^{-7}$	$3.61 \cdot 10^{-6}$
FH2001	0.031	$3.23 \cdot 10^{-8}$	$3.38 \cdot 10^{-7}$
FH2328	0.065	$1.81 \cdot 10^{-9}$	$3.04 \cdot 10^{-8}$
FH2004	0.090	$8.97 \cdot 10^{-11}$	$2.96 \cdot 10^{-9}$
CPH12	0.102	$1.26 \cdot 10^{-11}$	$6.27 \cdot 10^{-10}$
FH2010	0.070	$1.46 \cdot 10^{-12}$	$9.84 \cdot 10^{-11}$
PEZ17	0.080	$1.05 \cdot 10^{-13}$	$1.16 \cdot 10^{-11}$
PEZ16	0.170	$6.07 \cdot 10^{-15}$	$1.91 \cdot 10^{-12}$
CPH4	0.173	$8.54 \cdot 10^{-16}$	$4.92 \cdot 10^{-13}$
FH2096	0.131	$2.25 \cdot 10^{-16}$	$1.76 \cdot 10^{-13}$
vWF.X	0.192	$3.39 \cdot 10^{-17}$	$4.70 \cdot 10^{-14}$

Note. The loci are listed in the order of increasing F_{ST} . * Here and below the product of the genotype frequency of the previous and current loci; ** here and below the product of the genotype frequency of the previous and current loci.

The analysis of PIC measures by Botstein et al. (1980) revealed that all selected loci in the total sample of dogs are highly informative for the study of the DNA of both the domestic dog and the Grey wolf (Table 4).

In the combined sample of dogs, the minimum PIC value (0.472) was found at locus FH2096. The maximum PIC values were observed at FH2016 for the wolf (0.882) and at FH2016 for the dog (0.885). The average PIC values were 0.720 for the dog and 0.742 for the wolf, which can be considered significant for interpretation of results in a forensic genetic investigation.

While the allele sequences of the wolf and dog were identical, allele sequencing revealed the presence of simple repeats. Incomplete tandem repeats were identified in the alleles of the FH2016, FH2361, and FH2328 loci. Specifically, for the FH2016 and FH2328 loci, incomplete tandems were detected both in the sample of wolves and in the sample of dogs. For the FH2361 locus, microvariants were identified only in the sample of dogs.

Sequencing of the alleles of the FH2001 locus produced an unexpected result (Fig. 4). A 6 bp insertion located in the non-tandem region of the locus was found in the combined

Table 4. Values of the polymorphism information content of the selected loci for the samples of the Grey wolf and the domestic dog

Loci	PIC	
	Domestic dog	Grey wolf
FH2096	0.472	0.552
vWFX	0.541	0.813
CPH4	0.566	0.817
FH2079	0.620	0.579
CPH12	0.651	0.580
FH2010	0.657	0.747
PEZ17	0.766	0.746
PEZ16	0.769	0.827
FH2001	0.774	0.701
FH2004	0.778	0.876
FH2054	0.801	0.759
FH2361	0.830	0.788
VGL3438	0.845	0.755
FH2328	0.848	0.713
FH2016	0.885	0.882
Average	0.720	0.742

Note. PIC – polymorphism information content of a locus.

sample of wolves and dogs. This insertion was observed only in long alleles (with 10 or more tandem repeats), and there were no alleles with 11 or more repeats that did not contain insertions. The FH2001 locus was not excluded from the final forensic panel, and special names were assigned to alleles with insertion (“10in”–“14in” alleles).

Based on the results, we created the Cplex test system, which contains 15 STR loci and two sex loci. We obtained the nucleotide sequences of all the identified alleles; these alleles

were identified in the tandem format for compatibility of the panel with the instrumentation. The sequences were deposited in the GenBank database (Table 5).

The developed test system was validated in accordance with the protocol of the Scientific Working Group on DNA Analysis Methods (R.V. Guideline, 2004), and it was tested on benchmark samples and on real forensic objects. These methodological developments are being successfully used in expert practice when investigating the facts of illegal hunting, cruelty to animals and other offenses in the Republic of Belarus.

Conclusion

In this study, we selected 15 microsatellite loci and two sex loci that are suitable for forensic DNA examination of both the Grey wolf and the domestic dog. Furthermore, we designed a universal Cplex test system for the identification of individuals of the *C. lupus* species. Finally, we carried out an analysis of polymorphism and forensic parameters of loci and studied the genetic structure of the Belarusian populations of the *C. lupus* species. According to the results of statistical analysis of the dog and wolf genotype pools, the selected loci conform to the Hardy–Weinberg equilibrium. The coefficients of subdivision of the population for each STR locus of the test system were calculated, and the effectiveness of their use was proven. The developed Cplex test system was validated in accordance with the international standard and used in forensic research of cases of illegal hunting, animal attacks of people and livestock, as well as cases of cruelty to animals.

Based on the developed test system, we created two guides for practicing forensic experts, “The method of DNA-based identification of biological samples of the Grey wolf (*Canis lupus lupus*) and domestic dog (*Canis lupus familiaris*)” and “The method of using the computational tool for the analysis of genetic data of animals of the biological species *Canis lupus* – the Grey wolf (*Canis lupus lupus*) and the domestic dog (*Canis lupus familiaris*)”. The computational tool contains arrays of genotypes and a mathematical apparatus that allows one to automate the analysis of data when identifying species of the Grey wolf or the domestic dog, calculating the reliability of an expert conclusion in an identification study or establishing biological relationship; it also allows conducting DNA fingerprinting registration of biological samples of

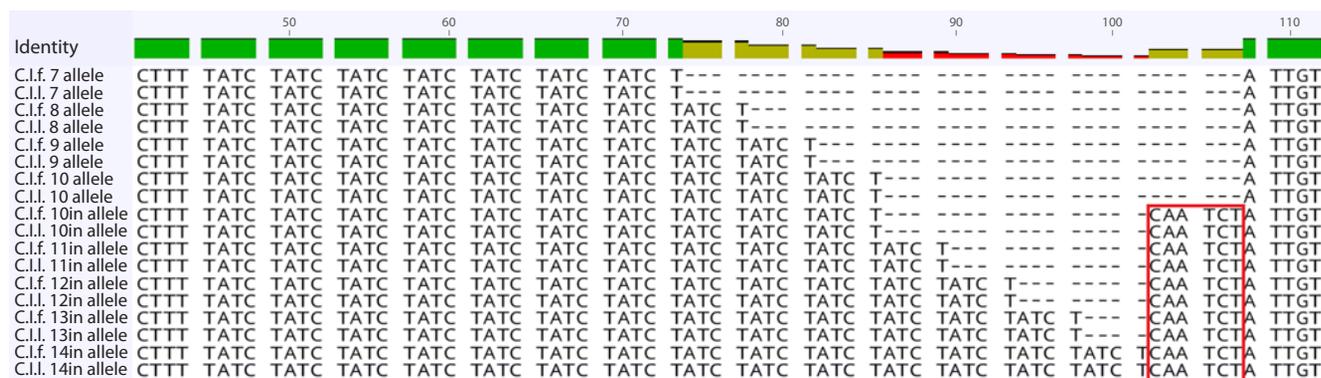


Fig. 4. Part of the sequence of the FH2001 locus in DNA of the dog and the wolf.

C.I.f. – domestic dog, C.I.l. – grey wolf. Red rectangle – 6 bp insertion in the non-tandem region of the locus.

Table 5. Characterization of microsatellite loci of the Cplex test system

Loci	Revealed range in bp	Tandem type and range	GenBank access number	
			Minimum length allele	Maximum length allele
DBX	246	–	OQ216490	
DBY	117	–	OQ216491	
FH2096	88–108	(AATG) _{5–10}	OQ216492	OQ216493
vWFX	133–199	(AGGAAT) _{5–16}	OQ216494	OQ216495
CPH4	138–152	(CA) _{14–22}	OQ216496	OQ216497
FH2079	260–292	(TGGGA) _{6–14}	OQ216498	OQ216499
CPH12	180–200	(AC) _{5–19}	OQ216500	OQ216501
FH2010	215–239	(GAAT) _{7–13}	OQ216502	OQ216503
PEZ17	190–222	(TTTC) _{11–19}	OQ216504	OQ216505
PEZ16	269–337	(GAAA) _{5–22}	OQ216506	OQ216507
FH2004	229–337	(TTCT) _{11–38}	OQ216508	OQ216509
FH2054	140–180	(ATCT) _{13–23}	OQ216510	OQ216511
VGL3438	101–145	(AAAG) _{10–20}	OQ216512	OQ216513
FH2001	124–158	(ATCT) _{7–10}	OQ216514	OQ216515
		(ATCT) _{10–14} CAACTC	OQ216516	OQ216517
FH2361	329–425	(TCTT) _{11–35}	OQ216518	OQ216519
		(TCTT) _{11–18} TC	OQ216520	OQ216521
FH2328	181–219	(AAAG) _{7–16}	OQ216522	OQ216523
		(AAAG) ₁₅ AA(AAAG) ₁	OQ216524	
FH2016	276–340	(CTTT) _{15–31}	OQ216525	OQ216526
		(CTTT) _{19–21} CT	OQ216527	OQ216528

animals of *C. lupus* species in a forensic examination. The developed methods are included in the Register of forensic methods and other methodological materials of the State Forensic Examination Committee of the Republic of Belarus, which constitutes the implementation of the development in the national legal system.

References

- Aspi J., Roininen E., Ruokonen M., Kojola I., Vila C. Genetic diversity, population structure, effective population size and demographic history of the Finnish wolf population. *Mol. Ecol.* 2006;15(6):1561–1576. DOI 10.1111/j.1365-294X.2006.02877.x
- Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Sayers E.W. GenBank. *Nucleic Acids Res.* 2005;34:D16–D20. DOI 10.1093/nar/gkj157
- Boitani L. Wolf Conservation and Recovery. Chicago: Univ. of Chicago Press, 2003;317–340
- Boom R.C.J.A., Sol C.J., Salimans M.M., Jansen C.L., Wertheim-van Dillen P.M., Van der Noordaa J.P.M.E. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* 1990;28(3):495–503. DOI 10.1128/jcm.28.3.495-503.1990
- Botstein D., White R.L., Skolnick M., Davis R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 1980;32(3):314
- Buckleton J.S., Curran J.M., Walsh S.J. How reliable is the sub-population model in DNA testimony? *Forensic Sci. Int.* 2006;157(2-3): 144–148. DOI 10.1016/j.forsciint.2005.04.004
- Buckleton J.S., Bright J.A., Taylor D. Forensic DNA Evidence Interpretation. CRC Press, 2016. DOI 10.4324/9781315371115
- Caniglia R., Fabbri E., Greco C., Galaverni M., Randi E. Forensic DNA against wildlife poaching: identification of a serial wolf killing in Italy. *Forensic Sci. Int. Genet.* 2010;4(5):334–338. DOI 10.1016/j.fsigen.2009.10.012
- Dan M., Šnjegota D., Veličković N., Stefanović M., Obreht Vidaković D., Čirović D. Genetic variability and population structure of grey wolf (*Canis lupus*) in Serbia. *Russ. J. Genet.* 2016;52(8):821–827. DOI 10.1134/S1022795416080044
- DeNise S., Johnston E., Halverson J., Marshall K., Rosenfeld D., McKenn S., Sharp T., Edwards J. Power of exclusion for parentage verification and probability of match for identity in American kennel club breeds using 17 canine microsatellite markers. *Anim. Genet.* 2004;35(1):14–17. DOI 10.1046/j.1365-2052.2003.01074.x
- Eichmann C., Berger B., Reinhold M., Lutz M., Parson W. Canine-specific STR typing of saliva traces on dog bite wounds. *Int. J. Leg. Med.* 2004;118(6):337–342. DOI 10.1007/s00414-004-0479-7
- Evanno G., Regnaut S., Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 2005;14(8):2611–2620. DOI 10.1111/j.1365-294X.2005.02553.x

- Excoffier L., Laval G., Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online*. 2005;1:47-50
- Fan Z., Silva P., Gronau I., Wang S., Armero A.S., Schweizer R.M., Ramirez O., Pollinger J., Galaverni M., Ortega Del-Vecchyo D., Du L., Zhang W., Zhang Z., Xing J., Vilà C., Marques-Bonet T., Godinho R., Yue B., Wayne R.K. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res*. 2016;26(2):163-173. DOI 10.1101/gr.197517.115
- Freedman A.H., Wayne R.K. Deciphering the origin of dogs: from fossils to genomes. *Annu. Rev. Anim. Biosci.* 2017;5:281-307. DOI 10.1146/annurev-animal-022114-110937
- Galinskayaa T.V., Schepetov D.M., Lysenkov S.N. Prejudices against microsatellite studies and how to resist them. *Russ. J. Genet.* 2019; 55(6):617-632. DOI 10.1134/S0016675819060043
- Godinho R., Llaneza L., Blanco J.C., Lopes S., Alvares F., Garcia E.J., Palacios V., Cortés Y., Tategón J., Ferrand N. Genetic evidence for multiple events of hybridization between wolves and domestic dogs in the Iberian Peninsula. *Mol. Ecol.* 2011;20(24):5154-5166. DOI 10.1111/j.1365-294x.2011.05345.x
- Hall T., Biosciences I., Carlsbad C.J.G.B.B. BioEdit: an important software for molecular biology. *GERF Bull. Biosci.* 2011;2(1):60-61
- Halverson J., Basten C. A PCR multiplex and database for forensic DNA identification of dogs. *J. Forensic Sci.* 2005;50(2):352-363
- Hedrick P.W. Genetics of Populations. Boston: Jones and Bartlett, 2000
- Kalinowski S.T., Taper M.L., Marshall T.C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 2007;16(5):1099-1106. DOI 10.1111/j.1365-294x.2007.03089.x
- Korablev M.P., Korablev N.P., Korablev P.N. Genetic diversity and population structure of the grey wolf (*Canis lupus* Linnaeus, 1758) and evidence of wolf × dog hybridisation in the centre of European Russia. *Mamm. Biol.* 2021;101(1):91-104. DOI 10.1007/s42991-020-00074-2
- Magory Cohen T., King R., Dolev A., Boldo A., Lichter-Peled A., Kahila Bar-Gal G. Genetic characterization of populations of the golden jackal and the red fox in Israel. *Conserv. Genet.* 2013;14(1):55-63. DOI 10.1007/s10592-012-0423-1
- Ministry of Forestry of the Republic of Belarus. Annual Forest Management Report. 2021 (in Russian)
- Parker H.G., Dreger D.L., Rimbault M., Davis B.W., Mullen A.B., Carpintero-Ramirez G., Ostrander E.A. Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Rep.* 2017;19(4):697-708. DOI 10.1016/j.celrep.2017.03.079
- Peakall R.O.D., Smouse P.E. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes.* 2006;6(1):288-295. DOI 10.1111/j.1471-8286.2005.01155.x
- Peakall R., Smouse P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics.* 2012;28:2537-2539. DOI 10.1093/bioinformatics/bts460
- Pritchard J.K., Stephens M., Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155(2):945-959. DOI 10.1093/genetics/155.2.945
- Radko A., Podbielska A. Microsatellite DNA analysis of genetic diversity and parentage testing in the popular dog breeds in Poland. *Genes.* 2021;12(4):485. DOI 10.3390/genes12040485
- Revised Validation Guideline – Scientific Working Group on DNA Analysis Methods (SWGDM). *Forensic Sci. Commun.* 2004;6(3). Available from: http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2004/standards/2004_03_standards02.htm
- Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 1977;74(12):5463-5467. DOI 10.1073/pnas.74.12.5463
- Sastre N., Vila C., Salinas M., Bologov V.V., Urios V., Sánchez A., Francino O., Ramírez O. Signatures of demographic bottlenecks in European wolf populations. *Conserv. Genet.* 2011;12(3):701-712. DOI 10.1007/s10592-010-0177-6
- Seddon J.M. Canid-specific primers for molecular sexing using tissue or non-invasive samples. *Conserv. Genet.* 2005;6(1):147-149. DOI 10.1007/s10592-004-7734-9
- Stronen A.V., Aspi J., Caniglia R., Fabbri E., Galaverni M., Godinho R., Kvist L., Mattucci F., Nowak C., von Thaden A., Harmoinen J. Wolf-dog admixture highlights the need for methodological standards and multidisciplinary cooperation for effective governance of wild × domestic hybrids. *Biol. Conserv.* 2022;266:109467. DOI 10.1016/j.biocon.2022.109467
- The Evaluation of Forensic DNA Evidence / Committee on DNA Forensic Science: an Update. National Research Council. Washington D.C.: National Academy Press, 1996
- van Asch B., Alves C., Santos L., Pinheiro R., Pereira F., Gusmao L., Amorim A. Genetic profiles and sex identification of found-dead wolves determined by the use of an 11-loci PCR multiplex. *Forensic Sci. Int. Genet.* 2010;4(2):68-72. DOI 10.1016/j.fsigen.2009.05.003
- van Oosterhout C., Hutchinson W.F., Wills D.P., Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes.* 2004;4(3):535-538. DOI 10.1111/j.1471-8286.2004.00684.x
- Verardi A., Lucchini V., Randi E. Detecting introgressive hybridization between free-ranging domestic dogs and wild wolves (*Canis lupus*) by admixture linkage disequilibrium analysis. *Mol. Ecol.* 2006; 15(10):2845-2855. DOI 10.1111/j.1365-294X.2006.02995.x
- Wright S. Variability Within and Among Natural Populations. Evolution and the Genetics of Populations. Chicago: Univ. of Chicago Press, 1978

ORCID

A.E. Hrebianchuk orcid.org/0000-0002-1224-3275
I.S. Tsybovsky orcid.org/0000-0002-8611-8215

Acknowledgements. The authors express their gratitude to the Republican State Public Association “Belarusian Society of Hunters and Fishermen”, veterinary center “Drug” and to the State Institution “Minsk city veterinary station” for their help in procuring a collection of biological samples.

Funding information. This research was carried out at Scientific and Practical Center of the State Forensic Examination Committee of the Republic of Belarus and was funded by the State Committee on Science and Technology of the Republic of Belarus (Project No. 20190195).

Conflict of interest. The authors declare no conflict of interest.

Received January 24, 2023. Revised July 4, 2023. Accepted July 11, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Comparative peculiarities of genomic diversity in *Gallus gallus domesticus* chickens with decorative plumage: the muffs and beard phenotype

N.V. Dementieva, Y.S. Shcherbakov, A.E. Ryabova, A.B. Vakhrameev, A.V. Makarova, O.A. Nikolaeva, A.P. Dysin, A.I. Azovtseva, N.R. Reinbah, O.V. Mitrofanova 

Russian Research Institute of Farm Animal Genetics and Breeding – Branch of the L.K. Ernst Federal Research Center for Animal Husbandry, Tyarlevo, St. Petersburg, Russia
 mo1969@mail.ru

Abstract. Throughout history, humans have been attempting to develop the ornamental features of domestic animals in addition to their productive qualities. Many chicken breeds have developed tufts of elongated feathers that jut out from the sides and bottom of the beak, leading to the phenotype known as muffs and beard. It is an incomplete autosomal dominant phenotype determined by the Mb locus localised on chromosome GGA27. This project aimed to analyse the genetic diversity of chicken breeds using full genomic genotyping with the Chicken 60K BeadChip. A total of 53,313 Single Nucleotide Polymorphisms were analysed. DNA was obtained from breeds with the muffs and beard as a marker phenotype: Faverolles ($n = 20$), Ukrainian Muffed ($n = 18$), Orloff ($n = 20$), Novopavlov White ($n = 20$), and Novopavlov Coloured ($n = 15$). The Russian White ($n = 20$) was selected as an alternative breed without the muffs and beard phenotype. The chickens are owned by the Centre of Collective Use “Genetic Collection of Rare and Endangered Breeds of Chickens” (St. Petersburg region, Pushkin), and are also included in the Core Shared Research Facility (CSRF) and/or Large-Scale Research Facility (LSRF). Multidimensional scaling revealed that the Novopavlov White and the Novopavlov Coloured populations formed a separate group. The Ukrainian Muffed and the Orloff have also been combined into a separate group. Based on cluster analysis, with the cross-validation error and the most probable number of clusters $K = 4$ taken into account, the Orloff was singled out as a separate group. The Ukrainian Muffed exhibited a notable similarity with the Orloff under the same conditions. At $K = 5$, the populations of the Novopavlov White and the Novopavlov Coloured diverged. Only at $K = 6$, a distinct and separate cluster was formed by the Ukrainian Muffed. The Russian White had the greatest number of short (1–2 Mb) homozygous regions. If the *HOXB8* gene is located between 3.402 and 3.404 Mb on chromosome GGA27, homozygous regions are rarely found in the chickens with the muffs and beard phenotype. Scanning the chicken genome with the Chicken 60K BeadChip provided enough information about the genetic diversity of the chicken breeds for the peculiarities of the development of the ornamental muffs and beard phenotypes in them to be understood. For example, Phoenix bantams, whose tail feathers grow throughout their lives, require greater consideration of husbandry conditions.

Key words: whole-genome genotyping; SNP marker; phenotype; genotype; genetic diversity; polymorphism; heterozygosity; DNA; chicken breeds.

For citation: Dementieva N.V., Shcherbakov Y.S., Ryabova A.E., Vakhrameev A.B., Makarova A.V., Nikolaeva O.A., Dysin A.P., Azovtseva A.I., Reinbah N.R., Mitrofanova O.V. Comparative peculiarities of genomic diversity in *Gallus gallus domesticus* chickens with decorative plumage: the muffs and beard phenotype. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(1):108-116. DOI 10.18699/vjgb-24-13

Сравнительные особенности геномного разнообразия кур *Gallus gallus domesticus* с декоративным фенотипом оперения «баки и борода»

Н.В. Дементьева, Ю.С. Щербаков, А.Е. Рябова, А.Б. Вахрамеев, А.В. Макарова, О.А. Николаева, А.П. Дысин, А.И. Азовцева, Н.Р. Рейнбах, О.В. Митрофанова 

Всероссийский научно-исследовательский институт генетики и разведения сельскохозяйственных животных – филиал Федерального исследовательского центра животноводства – ВИЖ им. академика Л.К. Эрнста, пос. Тярлево, Санкт-Петербург, Россия
 mo1969@mail.ru

Аннотация. На протяжении истории взаимодействия с домашними животными человек стремился усилить не только их продуктивные качества, но и различные декоративные особенности. У кур ряда пород сформировались пучки удлинённых перьев, выступающих сбоку и снизу от клюва, образуя фенотип, описываемый как «баки и борода» (англ. muffs and beard). Это неполный аутосомно-доминантный фенотип, кодируемый локусом Mb, ло-

кализованным на хромосоме GGA27. Цель нашей работы – проанализировать генетическое разнообразие пород кур, определенное с помощью полногеномного генотипирования с использованием чипов Chicken 60K BeadChip. Всего в анализе учитывалось 53 313 однонуклеотидных полиморфных замен (SNP). ДНК получена от пород, обладающих маркерным признаком «баки и борода»: фавероль ($n = 20$), украинская ушанка ($n = 18$), орловская ($n = 20$), новопавловская белая ($n = 20$) и новопавловская цветная ($n = 15$). В качестве альтернативной породы, не имеющей фенотипа «баки и борода», использовалась русская белая ($n = 20$). Птица содержалась в ЦКП «Генетическая коллекция редких и исчезающих пород кур» (г. Санкт-Петербург, Пушкин), входящем в состав Сетевой биоресурсной коллекции животных и птиц. Методом многомерного шкалирования установлено, что отдельную группировку образовали популяции новопавловская белая и новопавловская цветная. В самостоятельную группу объединились породы украинская ушанка и орловская. С помощью кластерного анализа с учетом ошибки кросс-валидации и наиболее вероятным числом кластеров $K = 4$ орловская порода выделена в отдельную группу. Украинская ушанка в этом случае продемонстрировала значительное сходство с орловской породой. При $K = 5$ разделились новопавловская белая и новопавловская цветная популяции. И только при $K = 6$ явный отдельный кластер образовала украинская ушанка. У кур русской белой породы отмечено наибольшее количество коротких (1–2 Мб) гомозиготных районов. В месте расположения гена *HOXB8* в регионе 3.402–3.404 Мб на хромосоме GGA27 у представителей пород с фенотипом «баки и борода» гомозиготные районы встречаются редко. Сканирование генома кур с использованием чипа Chicken 60K BeadChip позволяет получить достаточно информации о генетическом разнообразии пород кур для понимания особенностей формирования у них декоративного фенотипа «баки и борода». Ключевые слова: полногеномное генотипирование; SNP-маркер; фенотип; генотип; генетическое разнообразие; полиморфизм; гетерозиготность; ДНК; породы кур.

Introduction

The domestic chicken (*Gallus gallus domesticus*) is one of the most widespread domesticated animals in the world. This species assumes a prominent role in human society, providing the largest source of animal protein, as well as being an important factor in sociocultural development (Lawal, Hanotte, 2021). Since becoming domesticated, chickens have spread to different countries and continents, resulting in the numerous breeds we know today.

Genetic variability is a key part of the study of evolution, development and differentiation of living organisms. In domestic animals, breeds are organised into a specific system that evolves in accordance with the tasks defined by humans. As a result, there are remarkable phenotypes that distinguish domesticated animals from their wild ancestors.

Throughout history, humans have sought to improve not only the production of animals, but their ornamentation as well. Such features, when severe, can have a negative impact on the lives of individuals who possess them. However, a significant proportion of the ornamental traits common to different breeds of chickens have no negative effect on the animals. Some chicken breeds may exhibit morphological traits characterized by elongated feathers growing from the sides and bottom of the beak, creating a distinctive phenotype called “muffs and beard”. It is an incomplete autosomal dominant phenotype encoded by the Mb locus.

The feathers surrounding the beard and muffs in chickens vary considerably in shape and length. Certain breeds have a voluminous muff from ear to ear (Novopavlov chickens, Houndan, Crèvecoeur, etc.), others have a weak expression of it. In certain instances, such as in the case of Barbu d’Anvers chickens, it is mainly the muffs that are developed, and the throat part of the beard is barely visible.

When studying the genomic characteristics of chickens with the “muffs and beard” phenotype, it was proposed that the Mb allele is localised on GGA27, forming a complex structural variation in the genome that leads to altered expression of the *HOXB8* gene (Guo Y. et al., 2016). Other researchers

have also found regions implicated in the development of this trait by using whole-genome analysis of GGA1, GGA2 and GGA27. The analysis of the *HOXB8* gene family members showed that it had different evolutionary dynamics among animals and that its motifs were conserved among avians, reptiles, amphibians and mammals, except for fish, whose *HOXB8* protein lost motif 10. This suggests a potential role for *HOXB8* in the evolution of sophisticated skin structures such as keratinous appendages. The authors highlight the intricate protein interactions of the *HOXB* family gene products in chickens, which are thought to contribute to understanding of the mechanisms of development and differentiation of the “muffs and beard” phenotype (Yang et al., 2020).

There is currently limited information concerning the regulation of head feathering in chickens due to the phenotypic diversity of the muff and beard traits. Of particular significance is the necessity for employing gene pool breeds as model objects in the search for new candidate genes, for example ones associated with hair growth in humans and animals. Selection without a strict focus on productive traits enables the attainment of a greater genetic diversity in the gene pool of chicken breeds, unlike that of industrial populations. Therefore, the study of their genomes can provide new information about structural changes in genes, the accumulation of homozygous regions and other genetic features.

Genetic Collection of Rare and Endangered Chicken Breeds, which is a part of Core Shared Research Facility (CSRF) and/or Large-Scale Research Facility (LSRF), contains several breeds with this trait: Faverolles, Ukrainian Muffed, Orloff, and Novopavlov (Paronyan et al., 2016). Orloff Mille Fleur is a historic Russian chicken breed with distinct physical characteristics and a manifest “muffs and beard” trait. The Ukrainian Muffed belongs to local heritage breeds of the southern regions of Ukraine and Russia, and has a variety of plumage colours. Faverolles, a historic French breed, is renowned for its rich and delectable meat. Additionally, it possesses voluminous ornamental muffs and beard. The Novopavlov breed is a phenotypically restored ancient ornamental



Fig. 1. Expression of the “muffs and beard” phenotype in chickens of different breeds: *a* – Orloff; *b* – Faverolles; *c* – Novopavlov; *d* – Ukrainian Muffed.

breed of Pavlovo chickens, distinguished by their muffs and beard (Fig. 1). Furthermore, the Centre of Collective Use also includes breeds that do not have the Mb allele in their genome.

The purpose of our study was to examine the genetic variation of chicken breeds harboring the genetic trait “muffs and beard” via whole-genome genotyping with the Chicken 60K BeadChip, in order to obtain new information on structural changes in the genomes, accumulation of homozygous regions and other genetic characteristics.

Materials and methods

The material of the study was genomic DNA of chickens held in the bioresource collection of the Russian Research Institute of Farm Animal Genetics and Breeding – Branch of the L.K. Ernst Federal Research Centre for Animal Husbandry “Genetic Collection of Rare and Endangered Breeds of Chickens” (St. Petersburg-Pushkin), which is part of the Core Shared Research Facility (CSRF) and/or Large-Scale Research Facility (LSRF).

For this study, chickens were selected with the marker traits “muffs and beard”. These were Faverolles ($n = 20$), Ukrainian Muffed ($n = 18$), Orloff ($n = 20$), along with representatives from two groups of Novopavlov chickens: Novopavlov White ($n = 20$) and Novopavlov Coloured ($n = 15$). As an alternative breed without the “muffs and beard” phenotype, the Russian White breed was chosen ($n = 20$).

Chicken blood was collected from the axillary vein into microtubes containing 30 μ l 0.5 M EDTA anticoagulant. Genomic DNA was extracted by phenol-chloroform extraction.

To assess the purity of isolated DNA, its quality was determined by measuring the absorbance at 260 and 280 nm ($OD_{260/280}$) using a NanoDrop 2000 instrument (Thermo Fisher Scientific Inc., Waltham, MA, USA) in accordance with the manufacturer’s instructions. DNA samples with $OD_{260/280}$ values between 1.6 and 2.0 were chosen for whole-genome genotyping.

Whole-genome genotyping was performed with a medium density Chicken 60K BeadChip (Illumina Inc., USA) DNA chip containing ~50,000 SNPs. The acquired whole-genome data facilitated the evaluation of genetic diversity through DNA sequence polymorphism analysis. A comprehensive analysis of 53,313 SNPs was conducted.

Quality control and filtering of genotyping data for each SNP and each sample were carried out using the PLINK 1.9 software package (<http://zzz.bwh.harvard.edu/plink>). Sample

genotyping quality was assessed using the following filters: the proportion of genotyped SNPs out of the total number of SNPs on the DNA chip for each tested SNP in an individual sample was greater than 90 %; the genotyping quality for each tested SNP across all genotyped samples was also greater than 90 %; the frequency of minor allele occurrence was greater than 1 %; and the deviation of SNP genotypes from the Hardy–Weinberg equilibrium probability was less than 10^{-6} .

To evaluate the genetic diversity, the following indices were calculated using the R package *diveRsity* (Keenan et al., 2013): H_O for observed heterozygosity, H_E for expected heterozygosity, F for inbreeding coefficient, and F_{min} and F_{max} for minimum and maximum detected inbreeding coefficients, respectively.

In order to assess the structure of the genome of the various breeds, we used the method of analysing the number and average length of runs of homozygosity (ROHs), the method of multidimensional scaling (MDS) based on the identity-by-state (IBS) matrix, and the method of F_{ST} analysis of the genetic divergence of populations.

A method for detecting ROHs was employed using sequential SNP detection, through the R package *detectRUNS* (Biscarini et al., 2018). To avoid underestimating the quantity of ROHs exceeding 8 Mb in length, we permitted only one SNP with an absent genotype and no more than one probable heterozygous genotype (Ferenčaković et al., 2013). To prevent common ROHs, the minimum ROH length was set as 1 Mb.

To minimise the number of false-positive results, the minimum number of SNPs was calculated as follows (Purfield et al., 2012):

$$l = \frac{\log_e \frac{\alpha}{n_s \cdot n_i}}{\log_e (1 - \overline{het})}$$

Where l is the minimum number of SNPs forming ROHs, n_s is the number of SNPs per individual, n_i is the number of genotyped individuals, α is the false-positive rate of ROHs (set to 0.05) and \overline{het} is the mean heterozygosity of the total SNPs within population. Here, a minimum of 23 SNPs was found.

Initially, the number and length of ROHs were determined for each individual, and then their mean values within each breed were computed. Additionally, the ROH-based genomic inbreeding coefficient (F_{ROH}) was calculated as the ratio of the sum of the length of all ROHs per animal to the total length of the autosomal genome. Then, the number of ROHs in the

genome of the breeds under study was determined by length classes: (1, 2), (2, 4), (4, 8), (8, 16) and >16 Mb. To establish the genome proportion that is overlapped by various ROH segments, we added up ROHs for categories that possess diverse minimum lengths (>1, >2, >4, >8, and >16 Mb).

MDS was conducted in PLINK 1.9 (Anderson et al., 2010), followed by plotting in the R package ggplot2. The fixation index F_{ST} was calculated using the EIGENSOFT 6.1.4 (Price et al., 2006) package with graphical representation performed in the SplitsTree software (Huson, Bryant, 2006).

We assessed the genetic structure of the analysed breeds in the Admixture 1.3 (Alexander et al., 2009) software and visually displayed it using the R package Pophelper (Francis, 2017). The optimal number of clusters (K) was determined by using the most likely number of ancestral clusters and calculating cross-validation error (CV error) values in the Admixture 1.3 software.

The phylogenetic tree of the studied chicken populations was constructed using the Neighbor-Net (Bryant, Moulton, 2004) method in the iTOL service (Letunic, Bork, 2021) based on pairwise genetic distances F_{ST} .

The pairwise genetic distances F_{ST} were estimated using the R package StaMPP (Pembleton et al., 2013). The significance of the obtained results (P) was calculated based on the analysis of 100 permutations.

Results

Indicators of genetic diversity in the studied breeds are presented in Table 1.

Table 1. Genetic diversity in the studied breeds

Breed	<i>n</i>	H_O	H_E	F	F_{max}	F_{min}
RB	20	0.285 ± 0.001	0.296 ± 0.001	0.037 ± 0.003	0.191	-0.096
F	20	0.321 ± 0.006	0.338 ± 0.001	0.051 ± 0.017	0.202	-0.087
UU	18	0.365 ± 0.005	0.364 ± 0.001	-0.002 ± 0.013	0.136	-0.082
O	20	0.355 ± 0.006	0.344 ± 0.001	-0.031 ± 0.017	0.215	-0.128
PB	20	0.344 ± 0.003	0.338 ± 0.001	-0.016 ± 0.009	0.057	-0.104
P	15	0.344 ± 0.007	0.342 ± 0.001	-0.007 ± 0.020	0.184	-0.084

Note. *n* – number of individuals in the analysis; H_O – observed heterozygosity; H_E – expected heterozygosity; F – inbreeding coefficient; F_{max} – maximum detected inbreeding coefficient; F_{min} – minimum detected inbreeding coefficient. At $\alpha = 0.05$, there is no statistically significant difference.

RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population).

Table 2. Pairwise genetic distances of F_{ST} between the studied chicken breeds

Index	RB	F	UU	O	PB
F	0.220	0			
UU	0.136	0.133	0		
O	0.176	0.180	0.075	0	
PB	0.198	0.214	0.137	0.174	0
P	0.187	0.206	0.128	0.165	0.037

Note. RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population). The statistical significance for all pairs of comparisons is set at $p < 0.0001$.

Pairwise genetic distances F_{ST} , as shown in Table 2, ranged from 0.037 (Novopavlov Coloured and Novopavlov White) to 0.220 (between Russian White and Faverolles). The F_{ST} value for the identified cluster consisting of Ukrainian Muffed and Orloff was 0.075. A visual representation of the F_{ST} genetic distances is given in Figure 2.

Figure 3 shows the results of the MDS analysis visualising the data. This approach makes it possible to analyse and visualise the points corresponding to the objects of interest in a way that minimises the distance between them. The objects are plotted in the diagram based on the selected principal component system. The results of MDS indicated that the Novopavlov White and Novopavlov Coloured populations formed a distinct cluster, whereas the Ukrainian Muffed and Orloff chicken breeds formed a separate cluster. There was no change in clustering when comparing the various components.

The CV error calculation in Admixture cluster analysis indicated that there are likely four clusters (K) in our sample. The cross-validation error was the lowest in this case (CV error = 0.54930). According to the admixture analysis (Fig. 4), the Orloff breed was represented as a separate cluster at K = 4. The Ukrainian Muffed comprises of a genetically and phenotypically identical population that demonstrates noteworthy similarities with the Orloff breed. At K = 5, the populations of white and coloured individuals in Novopavlov chickens were separated. At K = 6, a distinct cluster was formed by the Ukrainian Muffed breed, which contained individuals with similar genetic structures to the Orloff breed.

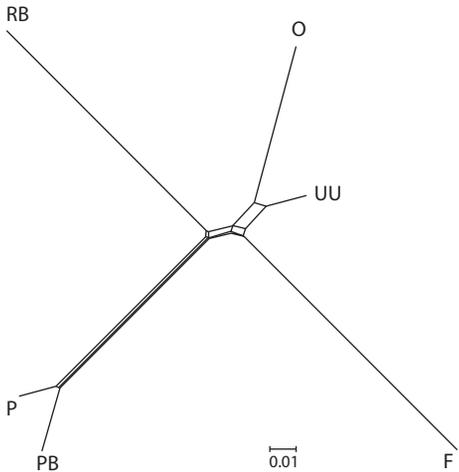


Fig. 2. Phylogenetic tree of the studied chicken populations based on pairwise genetic distances F_{ST} , constructed with the Neighbor-Net method.

RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population).

Analysis of the distribution of homozygous regions by length showed that the Russian White breed had the largest number of short homozygous regions (1–2 Mb), whereas the Faverolles breed had the smallest number of short homozygous regions (Fig. 5). Homozygous regions of class 16+ were more prevalent in the coloured population of the Novopavlov breed and the least prevalent in the Russian White.

Table 3 presents the descriptive statistics of the homozygous regions. The average number and mean length of extended homozygous regions were minimal in the investigated samples of the Ukrainian Muffed breed. On the contrary, the Faverolles breed showed the highest mean length and mean number of ROHs.

As the “muffs and beard” phenotype is linked to the *HOM8* gene situated on chromosome GGA27, we investigated the homozygous regions in this genetic segment separately. Figure 6 shows that ROH fragments are infrequent in breeds that exhibit the “muffs and beard” phenotype within the 3.402–3.404 Mb region on GGA27.

Discussion

Modern methods for studying genomic DNA polymorphism provide extensive data to comprehend the population’s entire genome architecture. Genotyping with chips of different densities allows the phylogenetic divergence of animal breeds to be assessed, and the information obtained can help to maintain distinct genetic diversity of popu-

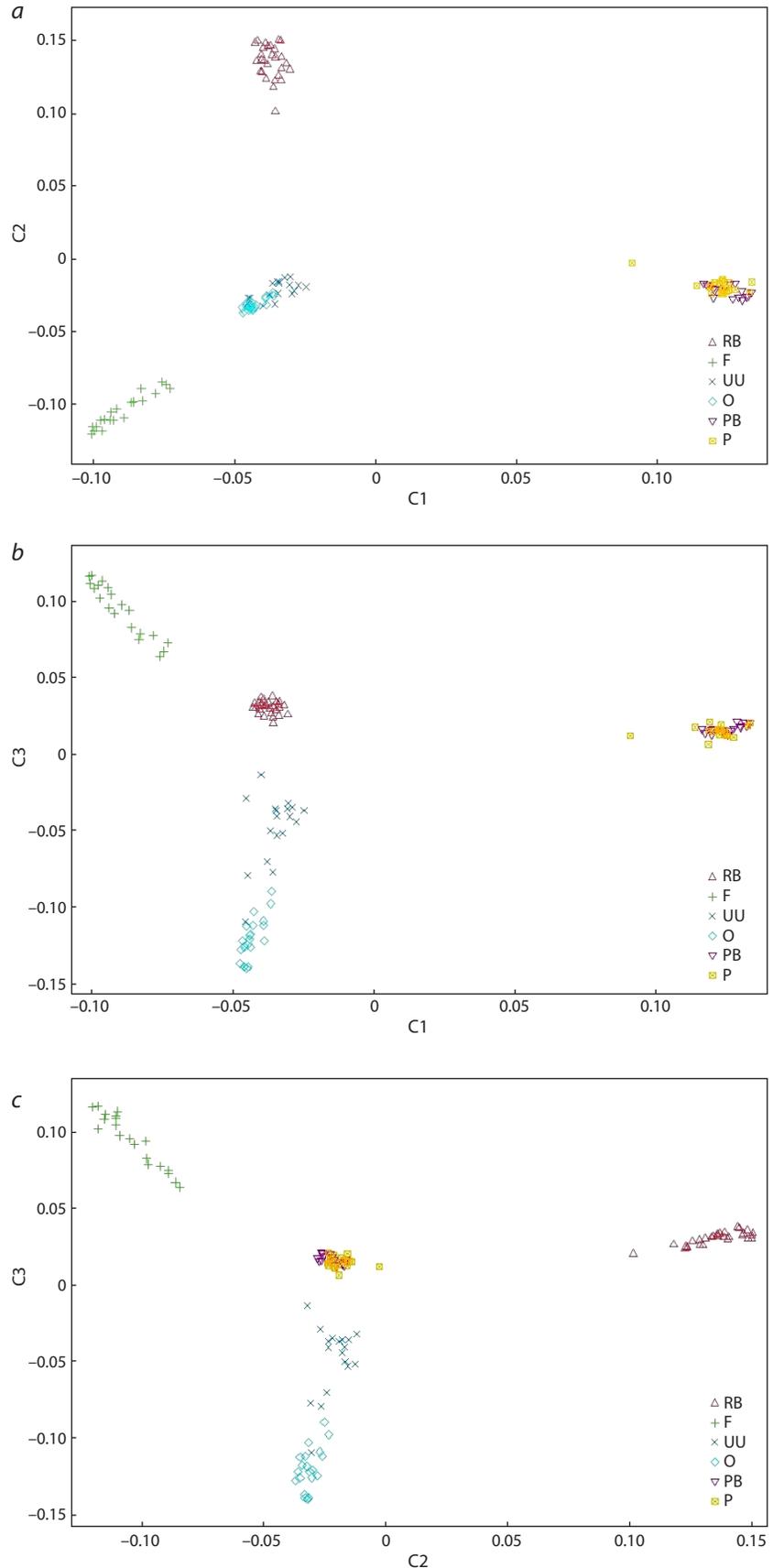


Fig. 3. Multidimensional scaling graph.

RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population).

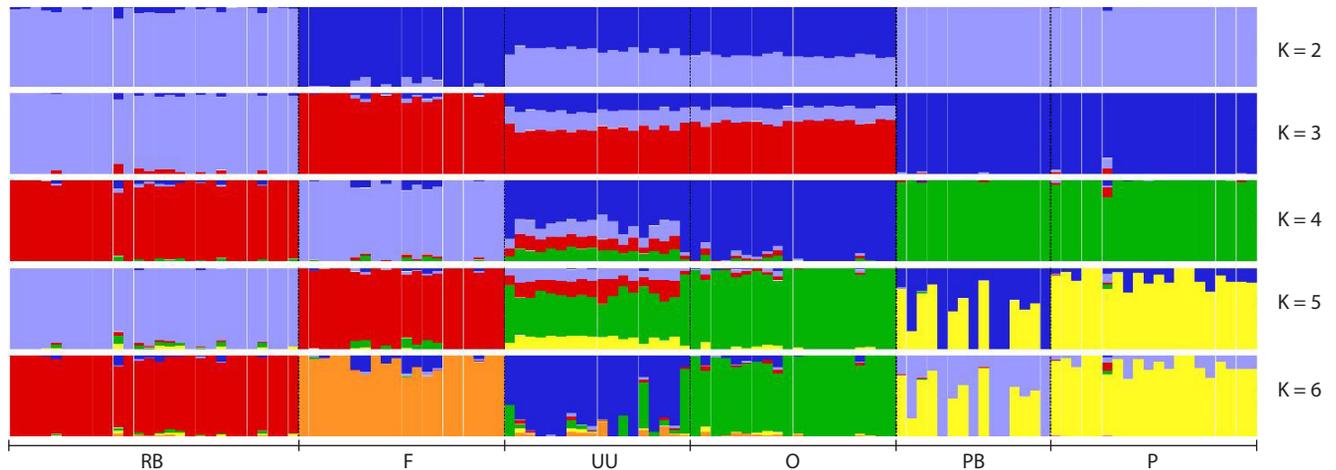


Fig. 4. Admixture cluster analysis conducted for six chicken breeds using whole-genome SNP analysis.
 RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population).

Table 3. Descriptive statistics of runs of homozygosity determined based on SNP genotypes in the studied chicken breeds

Population	n	Length of ROHs, Mb			Number of ROHs		
		X ± SE	Min	Max	X ± SE	Min	Max
RB	38	177.15 ± 5.29	126.12	234.06	143.71 ± 2.71	115	168
F	20	231.55 ± 11.51 ¹	142.65	313.17	155.50 ± 6.50 ²	109	206
UU	18	91.29 ± 9.20	48.20	200.17	78.11 ± 4.50	51	125
O	20	139.55 ± 10.01	89.60	291.26	105.00 ± 4.80	80	177
PB	15	224.42 ± 12.22 ²	166.15	341.79	153.07 ± 4.35 ²	133	192
P	20	184.90 ± 4.98	152.03	248.75	136.95 ± 2.60	108	153

Note. RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population).
^{1,2} Differences are not statistically significant. In other cases, differences are significant at a significance level of $\alpha = 0.05$.

lations (Dementieva et al., 2021; Krivoruchko et al., 2021). A comprehensive understanding of the genetic structures' specificity is crucial for examining genetic diversity, and it can be applied to explore the historical processes linked to the formation and evolution of populations as separate ecosystems influenced by human beings.

In this research, the application of the Chicken 60K Bead-Chip led to the identification of 53,313 single nucleotide polymorphisms. Previous genetic diversity study methods, based on the analysis of mini- and microsatellite loci and mitochondrial DNA, have much lower resolution (Fisinin et al., 2017; Guo H.W. et al., 2017).

The expected (H_E) and observed heterozygosity (H_O) (see Table 1) derived from the full genomic genotyping data in our study were higher in chickens that emerged as separate clusters in the MDS analysis (see Fig. 3). The initial group, which amalgamated the populations of the Novopavlov breed, exhibited H_O value of 0.344, while H_E values ranged from 0.338 ± 0.001 to 0.342 ± 0.001 . In the second group, comprised of the Ukrainian Muffed and Orloff breeds, the values for H_O were between 0.355 ± 0.006 and 0.365 ± 0.005 , and those for H_E varied between 0.344 ± 0.001 and 0.364 ± 0.001 (see Table 1). These findings concurred with other researchers' material (Strillacci et al., 2017; Yuan et al., 2022).

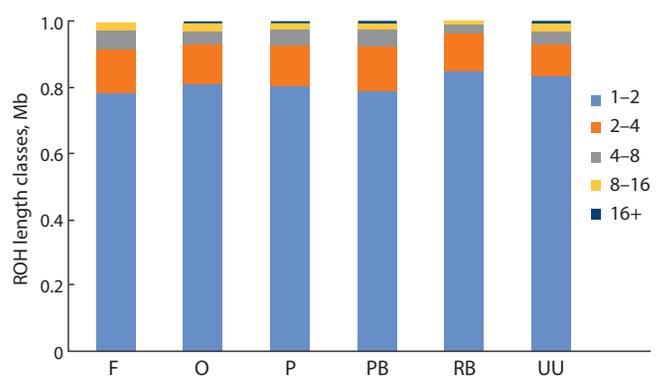


Fig. 5. Distribution of the number of runs of homozygosity (ROHs) in relation to their average length in the studied groups of chickens.
 RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population).

Based on the results of the multidimensional scaling analysis, it can be concluded that the breeds most distantly related to each other are the Russian White and Faverolles. The greatest genetic divergence between these breeds lies in their origins. The Russian White chicken breed was developed using white

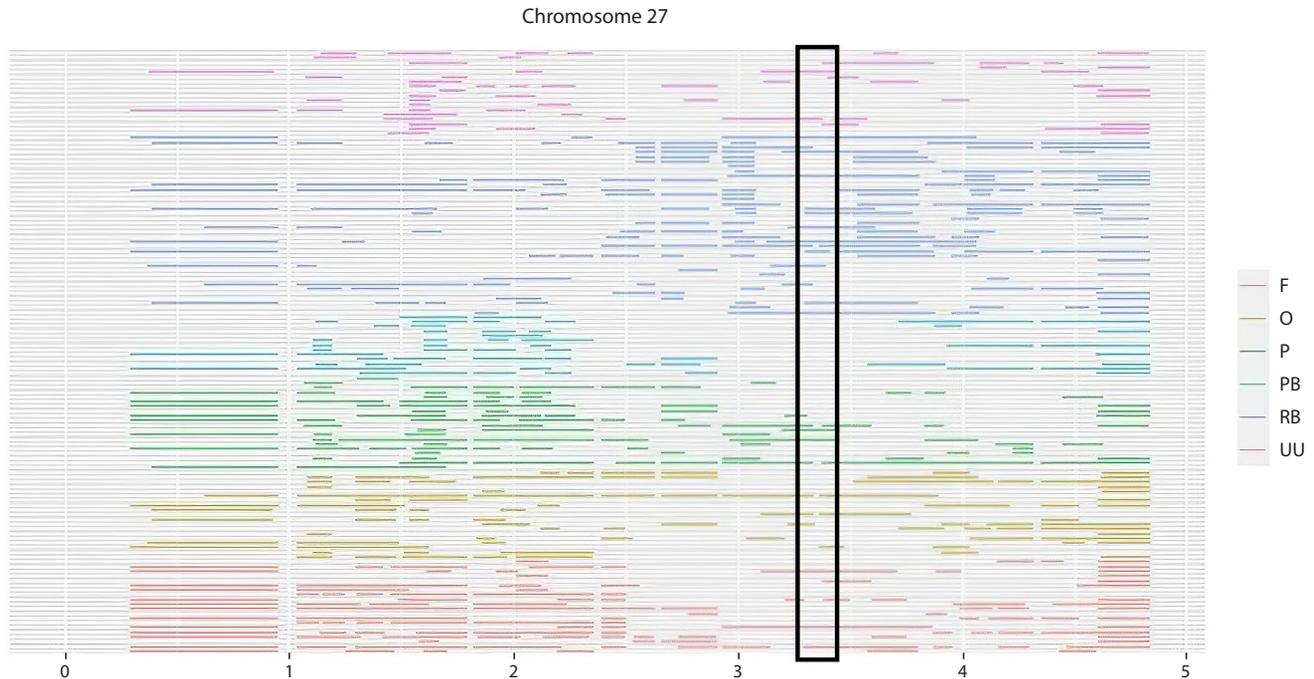


Fig. 6. Location of homozygosity runs on chromosome GGA27 in chickens.

The highlighted box indicates the region annotated with the Mb allele. RB – Russian White, F – Faverolles, UU – Ukrainian Muffed, O – Orloff, PB – Novopavlov (White population), P – Novopavlov (Coloured population).

domestic Leghorn chickens as a foundation (Dementieva et al., 2017). Faverolles were bred in France from indigenous chicken breeds. Breeds with similar genetics include Orloff Mille Fleur, Ukrainian Muffed and Faverolles. The White and Coloured populations of the Novopavlov breed are genetically related, as evidenced by their similar architecture. The Ukrainian Muffed breed has similarities to the Orloff Mille Fleur breed in genome fragments. This introgression of genomes between the breeds may have occurred during the late 19th and early 20th centuries, when both breeds flourished and developed in a common area. The utilization of diverse components during the analysis did not alter the spatial pattern of population arrangement, indicating that this method accurately reflects the genuine genetic divergence of breeds.

Assessment of genetic diversity provides greater insight into the genomic structure of chicken breeds and populations (Malomane et al., 2019; Restoux et al., 2022). Pairwise genetic distance F_{ST} visualization demonstrated a genetic correlation among the Orloff Mille Fleur, Ukrainian Muffed, and Faverolles breeds, as well as within the White and Coloured populations of the Novopavlov chicken breed (see Fig. 2).

The phylogenetic tree exhibits a conspicuous demarcation of the maximum divergence between the Faverolles and Russian White chicken breeds. The branch that contains the Faverolles, Ukrainian Muffed, and Orloff Mille Fleur breeds reveals their shared ancestry (see Fig. 2). Each mentioned breed forms its own branches in the future. The Orloff Mille Fleur chicken breed is dissimilar from the Ukrainian Muffed type, with contrasting traits and merging represented by the Faverolles chickens. The White and Coloured populations of the Novopavlov breed showed little divergence. Our findings

revealed a high level of genetic differentiation between breeds, comparable to literature data described previously in studies of other breeds (Dementieva et al., 2020; Fedorova et al., 2022).

Visualisation of the pairwise genetic distances of the F_{ST} using the Neighbourhood Network algorithm, as well as Admixture cluster analysis, confirmed the results of the multi-dimensional scaling analysis (see Fig. 4). At $K = 4$, the Orloff and Ukrainian Muffed populations were distinguished from the other populations. In addition, the F_{ST} analysis shows that the Ukrainian Muffed population under study has individuals that are genetically similar to the Orloff breed. This may result from the introgression of genomes between breeds. Unintended crossbreeding between Ukrainian Muffed and Orloff populations is possible due to the proximity of the breeding areas of these breeds in the past. At $K = 5$, the White and Coloured populations of the Novopavlov chicken breed were separated. These findings demonstrate that the populations possess an identical genetic structure, with the Novopavlov White strain having been acquired by selectively breeding white individuals from the coloured population.

The high frequency of brief homozygous regions (1–2 Mb) discovered within the Russian White and Ukrainian Muffed breeds suggests that long-term inbreeding has occurred (see Fig. 5). The Russian White chicken breed was obtained through rigorous selection for chick resistance to cold. As a result of a single crossing with White Leghorn in 2005, the Russian White population has no tendency to increase homozygosity, indicating high genetic diversity in the population. Long regions of homozygosity of class 16+ were greater in the Novopavlov chickens (White and Coloured populations), Orloff Mille Fleur, and Ukrainian Muffed breeds, indicating

the presence of recent inbreeding. The Russian White breed has the lowest number of long homozygous regions, an indication that there is no inbreeding within the population due to individual fixation of the producer during breeding (Fedorova et al., 2022; Mulim et al., 2022).

Location analysis of homozygous regions in the region annotated for the *HOXB8* gene responsible for the “muffs and beard” phenotype showed the absence of homozygous regions in breeds with this phenotype. This may be attributed to the region’s high variability. Guo’s study (Guo Y. et al., 2016) showed that the presence of the Mb allele causes the *HOXB8* gene to be ectopically expressed. As a result of the duplication of three regions on chromosome 27, a structural mutation takes place, which does not contribute to the selective accumulation of homozygous regions. ROH regions were discovered on chromosome GGA27, in the range of 1.5–1.6 Mb, occurring in more than 60 % of representatives of all breeds studied, except the Russian White, which lacks muffs and beard (see Fig. 6). Perhaps the buildup of homozygosity in these regions may not be a universal trait for all individuals in the investigated breeds. Consequently, it is plausible to assume that this genome region could serve as a marker for the “muffs and beard” characteristic with a certain degree of probability.

Conclusion

Based on the conducted study, we can state that full genomic genotyping using the Chicken 60K BeadChip (Illumina Inc., USA) medium-density DNA chips is an accurate method for analysing genetic divergence of chicken populations of different historical origins. Similar results can be achieved through a variety of statistical approaches to interpreting data. These results can be explained by considering the historical development of the breed ecosystem as well as the specific method used to detect polymorphism. The gathered data helps us comprehend the particularities of genomic architecture of the studied chicken breeds and populations, and to use this information to control variability in order to preserve genetic diversity. The findings obtained can be used in further research to identify candidate genes for the “muffs and beard” phenotype in chickens, as well as to use gene pool populations as model objects with a high level of genetic diversity.

References

- Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655-1664. DOI 10.1101/gr.094052.109
- Anderson C., Pettersson F., Clarke G., Cardon L.R., Morris A.P., Zondervan K.T. Data quality control in genetic case-control association studies. *Nat. Protoc.* 2010;5:1564-1573. DOI 10.1038/nprot.2010.116
- Biscarini F., Cozzi P., Orozco-ter Wengel P. Lessons learnt on the analysis of large sequence data in animal genomics. *Anim. Genet.* 2018;49(3):147-158. DOI 10.1111/age.12655
- Bryant D., Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* 2004;21(2): 255-265. DOI 10.1093/molbev/msh018
- Dementeva N.V., Romanov M.N., Kudinov A.A., Mitrofanova O.V., Stanishevskaya O.I., Terletsky V.P., Fedorova E.S., Nikitkina E.V., Plemashov K.V. Studying the structure of a gene pool population of the Russian White chicken breed by genome-wide SNP scan. *Selskokhozyaystvennaya Biologiya = Agricultural Biology.* 2017; 52(6):1166-1174. DOI 10.15389/agrobiol.2017.6.1166eng
- Dementieva N.V., Kudinov A.A., Larkina T.A., Mitrofanova O.V., Dysin A.P., Terletsky V.P., Tyshchenko V.I., Griffin D.K., Romanov M.N. Genetic variability in local and imported germplasm chicken populations as revealed by analyzing runs of homozygosity. *Animals (Basel).* 2020;10(10):1887. DOI 10.3390/ani10101887
- Dementieva N.V., Mitrofanova O.V., Dysin A.P., Kudinov A.A., Stanishevskaya O.I., Larkina T.A., Plemashov K.V., Griffin D.K., Romanov M.N., Smaragdov M.G. Assessing the effects of rare alleles and linkage disequilibrium on estimates of genetic diversity in the chicken populations. *Animal.* 2021;15(3):100171. DOI 10.1016/j.animal.2021.100171
- Fedorova E.S., Dementieva N.V., Shcherbakov Y.S., Stanishevskaya O.I. Identification of key candidate genes in runs of homozygosity of the genome of two chicken breeds, associated with cold adaptation. *Biology (Basel).* 2022;11(4):547. DOI 10.3390/biology11040547
- Ferenčaković M., Sölkner J., Curik I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 2013;45:42. DOI 10.1186/1297-9686-45-42
- Fisinin V.I., Selionova M.I., Shinkarenko L.A., Shcherbakova N.G. Study of microsatellites in the Russian breeds of turkey. *Selskokhozyaystvennaya Biologiya = Agricultural Biology.* 2017;52(4):739-748. DOI 10.15389/agrobiol.2017.4.739eng
- Francis R.M. Pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* 2017;17(1):27-32. DOI 10.1111/1755-0998.12509
- Guo H.W., Li C., Wang X.N., Li Z.J., Sun G.R., Li G.X., Liu X.J., Kang X.T., Han R.L. Genetic diversity of mtDNA D-loop sequences in four native Chinese chicken breeds. *Br. Poult. Sci.* 2017;58(5): 490-497. DOI 10.1080/00071668.2017.1332403
- Guo Y., Gu X., Sheng Z., Wang Y., Luo C., Liu R., Qu H., Shu D., Wen J., Crooijmans R.P., Carlborg Ö., Zhao Y., Hu X., Li N. A complex structural variation on chromosome 27 leads to the ectopic expression of *HOXB8* and the muffs and beard phenotype in chickens. *PLoS Genet.* 2016;12(6):e1006071. DOI 10.1371/journal.pgen.1006071
- Huson D., Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 2006;23(2):254-267. DOI 10.1093/molbev/msj030
- Keenan K., McGinnity P., Cross T.F., Crozier W.W., Prodöhl P.A. diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* 2013;4(8):782-788. DOI 10.1111/2041-210X.12067
- Krivoruchko A.Y., Skokova A.V., Yatsyk O.A., Kanibolotskaya A.A. Modern approaches to the genetic identification of farm animal breeds (review). *Agrarnaya Nauka Evro-Severo-Vostoka = Agricultural Science Euro-North-East.* 2021;22(3):317-328. DOI 10.30766/2072-9081.2021.22.3.317-328 (in Russian)
- Lawal R.A., Hanotte O. Domestic chicken diversity: origin, distribution, and adaptation. *Anim. Genet.* 2021;52(4):385-394. DOI 10.1111/age.13091
- Letunic I., Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021; 49(W1):W293-W296. DOI 10.1093/nar/gkab301
- Malomane D.K., Simianer H., Weigend A., Reimer C., Schmitt A.O., Weigend S. The SYNBREED chicken diversity panel: a global resource to assess chicken diversity at high genomic resolution. *BMC Genomics.* 2019;20(1):345. DOI 10.1186/s12864-019-5727-9
- Mulim H.A., Brito L.F., Pinto L.F.B., Ferraz J.B.S., Grigoletto L., Silva M.R., Pedrosa V.B. Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC Genomics.* 2022;23(1):209. DOI 10.1186/s12864-022-08384-0
- Paronyan I.A., Jurchenko O.P., Vakhrameev A.B., Makarova A.V. Breeding of indigenous and rare breeds of chickens. *Genetika i Razvedenie Zhivotnyh = Animal Genetics and Breeding.* 2016;4:62-66 (in Russian)

- Pembleton L.W., Cogan N.O., Forster J.W. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol. Ecol. Resour.* 2013;13(5):946-952. DOI 10.1111/1755-0998.12129
- Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A., Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 2006;38(8):904-909. DOI 10.1038/ng1847
- Purfield D.C., Berry D.P., McParland S., Bradley D.G. Runs of homozygosity and population history in cattle. *BMC Genet.* 2012;13:70. DOI 10.1186/1471-2156-13-70
- Restoux G., Rognon X., Vieaud A., Guemene D., Petitjean F., Rouger R., Brard-Fudulea S., Lubac-Paye S., Chiron G., Tixier-Boichard M. Managing genetic diversity in breeding programs of small populations: the case of French local chicken breeds. *Genet. Sel. Evol.* 2022;54:56. DOI 10.1186/s12711-022-00746-2
- Strillacci G., Cozzi M.C., Gorla E., Mosca F., Schiavini F., Román-Ponce S.I., Ruiz López F.J., Schiavone A., Marzoni M., Cerolini S., Bagnato A. Genomic and genetic variability of six chicken populations using single nucleotide polymorphism and copy number variants as markers. *Animal.* 2017;11(5):737-745. DOI 10.1017/S1751731116002135
- Yang K.X., Zhou H., Ding J.M., He C., Niu Q., Gu C.J., Zhou Z.X., Meng H., Huang Q.Z. Copy number variation in *HOXB7* and *HOXB8* involves in the formation of beard trait in chickens. *Anim. Genet.* 2020;51(6):958-963. DOI 10.1111/age.13011
- Yuan J., Li S., Sheng Z., Zhang M., Liu X., Yuan Z., Yang N., Chen J. Genome-wide run of homozygosity analysis reveals candidate genomic regions associated with environmental adaptations of Tibetan native chickens. *BMC Genomics.* 2022;23(1):91. DOI 10.1186/s12864-021-08280-z

ORCID

N.V. Dementieva orcid.org/0000-0003-0210-9344
Y.S. Shcherbakov orcid.org/0000-0001-6434-6287
A.E. Ryabova orcid.org/0000-0003-2362-2892
A.B. Vakhrameev orcid.org/0000-0001-5166-979X
A.V. Makarova orcid.org/0000-0002-3281-4581

O.A. Nikolaeva orcid.org/0000-0003-3828-1111
A.P. Dysin orcid.org/0000-0002-4468-0365
A.I. Azovtseva orcid.org/0000-0002-2963-378X
N.R. Reinbah orcid.org/0000-0001-6193-5617
O.V. Mitrofanova orcid.org/0000-0003-4702-2736

Acknowledgements. We are thankful for the support of this work provided by the Centre of Collective Use, the owner of Genetic Collection of Rare and Endangered Breeds of Chickens.

Funding statement. This work was financially supported by the Ministry of Science and Higher Education of the Russian Federation, Grant No. 075-15-2021-1037 (Internal No. 15.BRK.21.0001).

Conflict of interest. The authors declare no conflict of interest.

Received March 28, 2023. Revised September 25, 2023. Accepted September 26, 2023.

Original Russian text <https://vavilovj-icg.ru/>

Association of three single nucleotide polymorphisms in the *LPIN1* gene with milk production traits in cows of the Yaroslavl breed

A.V. Igoshin¹, T.M. Mishakova¹, R.B. Aitnazarov¹, A.V. Ilina², D.M. Larkin³, N.S. Yudin¹ 

¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Federal Williams Research Center for Forage Production and Agroecology, Scientific Research Institute of Livestock Breeding and Forage Production, Yaroslavl Region, Russia

³ Royal Veterinary College, University of London, London, United Kingdom

 yudin@bionet.nsc.ru

Abstract. Lipin-1 is a member of the evolutionarily conserved family of proteins and is expressed predominantly in adipose tissue and skeletal muscle. On the one hand, lipin-1 is an enzyme that catalyzes the dephosphorylation of phosphatidic acid to diacylglycerol (DAG) and thus participates in the metabolic pathways of biosynthesis of storage lipids in the cell, membrane phospholipids, and intracellular signaling molecules. On the other hand, lipin-1 is able to be transported from the cytoplasm to the nucleus and is a coactivator of lipid metabolism gene transcription. It was shown, using the analysis of single nucleotide polymorphism (SNP) associations, that the lipin-1 coding gene (*LPIN1*) is a promising candidate gene for milk production traits in Holstein and Brown Swiss cows. However, it is unclear how much of its effect depends on the breed. The Yaroslavl dairy cattle breed was created in the 18–19 centuries in Russia by breeding northern Great Russian cattle, which were short and poor productive, but well adapted to local climatic conditions and bad food base. It was shown by whole genome genotyping and sequencing that the Yaroslavl breed has unique genetics compared to Russian and other cattle breeds. The aim of the study was to assess the frequency of alleles and genotypes of three SNPs in the *LPIN1* gene and to study the association of these SNPs with milk production traits in Yaroslavl cows. Blood samples from 142 cows of the Yaroslavl breed were obtained from two farms in the Yaroslavl region. Genotyping of SNPs was carried out by polymerase chain reaction-restriction fragment length polymorphism method. Associations of SNPs with 305-day milk yield, fat yield, fat percentages, protein yield, and protein percentages were studied from the first to the fourth lactation. Statistical tests were carried out using a mixed linear model, taking into account the relationship between individuals. We identified three SNPs – rs110871255, rs207681322 and rs109039955 with a frequency of a rare allele of 0.042–0.261 in Yaroslavl cows. SNP rs110871255 was associated with fat yield during the third and fourth lactations. SNP rs207681322 was associated with milk yield for the second, third and fourth lactations, as well as protein yield for the third lactation. Thus, we identified significant associations of SNPs rs207681322 and rs110871255 in the *LPIN1* gene with a number of milk production traits during several lactations in Yaroslavl cows.

Key words: cow; Yaroslavl breed; milk yield; fat percentage; protein percentage; fat yield; protein yield; *LPIN1* gene; single nucleotide polymorphism; association.

For citation: Igoshin A.V., Mishakova T.M., Aitnazarov R.B., Ilina A.V., Larkin D.M., Yudin N.S. Association of three single nucleotide polymorphisms in the *LPIN1* gene with milk production traits in cows of the Yaroslavl breed. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov Journal of Genetics and Breeding*. 2024;28(1):117-125. DOI 10.18699/vjgb-24-14

Ассоциация трех однонуклеотидных полиморфизмов в гене *LPIN1* с показателями молочной продуктивности у коров ярославской породы

А.В. Игошин¹, Т.М. Мишакова¹, Р.Б. Айтназаров¹, А.В. Ильина², Д.М. Ларкин³, Н.С. Юдин¹ 

¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Ярославский научно-исследовательский институт животноводства и кормопроизводства – филиал Федерального научного центра кормопроизводства и агроэкологии им. В.Р. Вильямса, пос. Михайловский, Ярославская область, Россия

³ Королевский ветеринарный колледж, Лондон, Великобритания

 yudin@bionet.nsc.ru

Аннотация. Липин-1 является членом эволюционно-консервативного семейства белков и экспрессируется преимущественно в жировой ткани и скелетных мышцах. С одной стороны, липин-1 – это фермент, который дефосфорилирует фосфатидную кислоту до диацилглицерина и таким образом участвует в метаболических путях биосинтеза запасных липидов в клетке, фосфолипидов мембраны и внутриклеточных сигнальных молекул. С другой стороны, липин-1 способен транспортироваться из цитоплазмы в ядро и служит коактиватором транскрипции генов липид-

ного метаболизма. С использованием анализа ассоциаций однонуклеотидных полиморфизмов (ОНП) было показано, что ген липина-1 (*LPIN1*) является перспективным геном-кандидатом признаков молочной продуктивности у коров голштинской и бурой швицкой пород. Однако неясно, насколько его эффект зависит от породы. Ярославская молочная порода крупного рогатого скота была выведена в XVIII–XIX вв. в России путем разведения «в себе» северного великорусского скота, который был низкорослым и малопродуктивным, но хорошо адаптированным к местным климатическим условиям и скудной кормовой базе. С помощью полногеномного генотипирования и секвенирования было показано, что ярославская порода обладает уникальной генетикой по сравнению с российскими и зарубежными породами крупного рогатого скота. Целью работы была оценка частоты аллелей и генотипов трех ОНП в гене *LPIN1* и исследование ассоциации этих ОНП с показателями молочной продуктивности у коров ярославской породы. Образцы крови от 142 коров ярославской породы были получены из двух хозяйств Ярославской области. Генотипирование ОНП выполняли методом анализа полиморфизма длин рестрикционных фрагментов после проведения полимеразной цепной реакции. Ассоциации ОНП с удоем, выходом молочного жира и белка, а также с процентным содержанием жира и белка в молоке за 305 дней лактации были исследованы с первой по четвертую лактацию. Для статистического анализа использовали смешанную линейную модель с учетом родства между индивидами. При исследовании коров ярославской породы выявили три ОНП: rs110871255, rs207681322 и rs109039955 с частотой редкого аллеля 0.042–0.261. ОНП rs110871255 был ассоциирован с выходом жира за третью и четвертую лактации, rs207681322 был ассоциирован с удоем за вторую, третью и четвертую лактации, а также с выходом белка за третью лактацию. Таким образом, мы выявили достоверные ассоциации ОНП rs207681322 и rs110871255 в гене *LPIN1* с рядом показателей молочной продуктивности в ходе нескольких лактаций у коров ярославской породы. Ключевые слова: корова; ярославская порода; удой; процент жира; процент белка; выход жира; выход белка; ген *LPIN1*; однонуклеотидный полиморфизм; ассоциация.

Introduction

The most important economic trait in dairy farming is cow milk productivity which includes milk, fat and protein yields, and fat and protein percentage (Gutierrez-Reinoso et al., 2021). All of these characteristics are complex quantitative traits controlled by a large number of genes having little impact on phenotype (Weller et al., 2017; Silpa et al., 2021; Bekele et al., 2023; Singh et al., 2023). Recently, such methods as quantitative trait loci (QTL) mapping, genome-wide association studies (GWAS) (Bekele et al., 2023; Chen S.Y. et al., 2023; Teng et al., 2023), targeted RNA sequencing (RNA-seq) (Fang et al., 2020; Ahmad et al., 2021) and detecting signatures of selection in genomes (Rajawat et al., 2022; Nayak et al., 2023; Persichilli et al., 2023) have been widely applied to identify the genes and mutations directly affecting milk yield and milk composition (Khatkar et al., 2004; Weller, Ron, 2011; Lopdell, 2023).

Previously, when analyzing the whole-genome genotyping (WGG) data, we identified selection signatures on chromosome 11 in a group of European dairy and dual-purpose (Bestuzhev, Holstein, Kholmogory, Black Pied, and Yaroslavl) cattle breeds, and the top SNP was localized in the lipin-1 (*LPIN1*) gene (Yurchenko et al., 2018a). Other authors also found selection signatures in this gene when analyzing the WGG data of the Yaroslavl and Holstein breeds (Zinovieva et al., 2020). However, our study did not detect any selection signatures in the *LPIN1* gene region when analyzing the whole-genome sequencing data of the Yaroslavl breed (Ruvinskiy et al., 2022), which may be due to both a high significance threshold ($q = 0.01$) and a different set of breeds (Holstein, Kholmogory, Yakut) used for comparison.

Lipin-1 is a member of an evolutionarily conserved family that is represented by three proteins (lipins 1, 2, and 3) in most vertebrates (Csaki et al., 2013; Siniossoglou, 2013; Chen Y. et al., 2015; Saydakova et al., 2021). Lipin-1 is expressed predominantly in adipose tissues, skeletal muscles and, to a lesser extent, in the liver, brain and other tissues (Reue, Zhang,

2008). In one respect, it is a phosphohydrolase enzyme that dephosphorylates phosphatidic acid to diacylglycerol (DAG) and thus participates in the metabolic pathways for biosynthesis of cellular-storage lipids, membrane phospholipids, and intracellular signalling molecules. At the same time, lipin-1 is transported from the cytoplasm to the nucleus to coactivate lipid metabolism gene transcription. Although this protein lacks a DNA-binding domain, it has been shown to regulate transcription by interacting with other transcription factors (TF), e. g., it regulates adipocyte differentiation and functioning by interacting with the PPARgamma transcription factor (Kim et al., 2013), and fatty acid oxidation gene expression by interacting with the PPARalpha TF (Barroso et al., 2011). Also, lipin-1 binds to the mTORC1 protein complex and thus regulates the activity of the SREBP TF that, in turn, regulates multiple pathways for fatty acid, triglyceride and cholesterol biosynthesis (Peterson et al., 2011).

According to the NCBI Gene database, in cattle, the *LPIN1* gene spans about 136 Kb on chromosome 11, consists of 25 exons, and encodes eight transcripts translated into proteins ranging in size from 895 to 1,010 amino acids (<https://www.ncbi.nlm.nih.gov/gene/537224>). *LPIN1* mRNA expression in cow liver and mammary gland increases significantly at the peak of lactation compared to the beginning of lactation and dry periods (Bionaz, Looor, 2008; Li et al., 2020). Keeping lactating Holstein cows on a diet supplemented with fish and soybean oils to reduce milk fat yield caused an increase in *LPIN1* mRNA expression in their subcutaneous adipose tissue (Thering et al., 2009). *In vitro* treatment of bovine mammary gland cells with rosiglitazone (BRL49653), a PPARgamma-selective agonist, resulted in activation of *LPIN1* mRNA expression (Kadegowda et al., 2009).

The rs137457402 and rs136905033 SNPs in this gene were associated with the content of five fatty acids in the milk of Brown Swiss cows (Pegolo et al., 2016). The rs137457402 SNP was also associated with the percentage of protein content in the milk of the same cows (Cecchinato et al., 2014).



Yaroslavl cattle.

Han B. et al. (2019) showed the relationship of seven SNPs with milk yield, fat or protein percentage, and fat or protein yield in Chinese Holsteins, yet most of these associations were revealed only for the first or second lactations. Two nonsynonymous SNPs in the sixth exon of the *LPIN1* gene were associated with fat and protein percentages in the milk of Holstein-Friesian×Jersey crossbred cows from New Zealand (Du et al., 2021). All of the above indicates that *LPIN1* has been a promising candidate gene for milk productivity traits, but it is unclear to what extent its effect is breed-specific.

The Yaroslavl dairy breed was created in the 18–19th centuries in Russia in Yaroslavl Province (Dmitriev, 1978; Dmitriev, Ernst, 1989; Dunin, Dankvert, 2013; Stolpovsky et al., 2022). The animals are mostly black, with the head, belly, lower limbs and tip of the tail being white. They have characteristic black glasses-like markings around the eyes (see the Figure). The breed was created based on inter-se mating of Northern Great Russian cattle which was stunted and unproductive but well adapted to local climatic conditions and poor forage. The initial selection was based on the exterior and then on milk yield and fat percentage. In the 19–20th centuries, the Yaroslavl breed was crossbred with the Tyrolean, Angeln, Simmental, Algauz, Jersey, Dutch and Kholmogory cattle. In the USSR, crossbreeding with the Ostfriesian and Holstein bulls was also carried out. However, these crosses are believed to have had a small impact, as the Yaroslavl cattle have retained their specific exterior (Dmitriev, 1978; Dmitriev, Ernst, 1989; Stolpovsky et al., 2022).

In 2022, the total breed population was about 30,000 animals that are characterized by high milk yield (6,590 kg for 305 lactation days) and fat percentage (4.13 %) (Shichkin et al., 2023). WGG (Iso-Touru et al., 2016) and microsatellite

analysis (Abdelmanova et al., 2020) have demonstrated that the Yaroslavl breed has unique genetic parameters compared to Russian and foreign livestock, and foreign breeds have had a minor impact on the Yaroslavl cattle's gene pool (Sermyagin et al., 2018; Yurchenko et al., 2018b; Zinovieva et al., 2020).

The purpose of the present study was to estimate the allele and genotype frequencies of three SNPs in the *LPIN1* gene and to study the association of these SNPs with milk productivity traits in Yaroslavl cows.

Materials and methods

As a material for the study, blood samples from 142 Yaroslavl cows from two farms of the Yaroslavl Region were used. The phenotypic data extracted from breeding record cards were provided by the Selex Information and Analytical System.

DNA extraction was performed using the standard phenol-chloroform extraction method with preliminary proteolytic treatment (Sambrook, Russell, 2006). Genotyping of the rs110871255, rs109039955, and rs207681322 SNPs in the *LPIN1* gene was carried out by restriction fragment length polymorphism (RFLP) analysis after polymerase chain reaction (PCR). Primers were designed using the Vector NTI software package (Lu, Moriyama, 2004). The specificity of each primer pair was evaluated *in silico* using the primer-BLAST algorithm (Ye et al., 2012). The primers, PCR reaction conditions and restriction enzymes are given in Supplementary Material 1¹. The test for deviation from Hardy–Weinberg equilibrium and linkage disequilibrium between the studied SNPs (LD) were calculated in PLINK v1.9 (--ld option) (Purcell et

¹ Supplementary Materials 1 and 2 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Igoshin_Engl_28_1.pdf

al., 2007). For this purpose, the genotypic data were converted to the PED format recognized by the program.

The studied associations included those related to such traits as milk, milk fat and protein yields as well as milk fat and protein percentages for 305 lactation days. Data from four lactations were included in the analysis. If a cow's lactation period was less than 305 days, its milk, fat and protein yields for this period were standardized to 305 days by the formula (Wiggans, Van Vleck, 1979):

$$\hat{Y}_{305} = \left[1 + F_n \times \left(\frac{305-n}{n} \right) \right] \times Y_n.$$

Here, \hat{Y}_{305} is the expected productivity for 305 days; n is the actual lactation duration in days; Y_n is the productivity for the actual lactation period; F_n is Shook's factor for the n -th day to account for the decrease in productivity during lactation (Hillers, Williams, 1981). This factor is calculated based on the productivity data for the animals with lactation duration ≥ 305 days as

$$F_n = RY / [LP \times (305 - n)],$$

where RY is the difference between the 305-day productivity and that for n days; and LP is the productivity at the n -th day. The daily data required to calculate the corresponding values were derived from cumulative productivity curves (Supplementary Material 2). The protein and fat percentages for incomplete lactations were calculated from the data standardized to 305 days.

Statistical analysis was performed using the mixed linear model implemented in the "lme4qtl" R package ("relmatLmer" function) (Ziyatdinov et al., 2018). This model allows one to account for genetic relationships between individuals by modelling polygenic effects based on a kinship matrix. The matrix was calculated using the "kinship2" R package ("kinship" function) (Sinnwell et al., 2014), based on the animal pedigree data from their breeding record cards. Genotypes were coded as 0, 1, and 2 implying the additive contribution of SNP alleles to a trait. The cows' birth and calving (one preceding the lactation period) seasons were used as additional predictors. Since the error in calculating the expected milk yield for 305 days was probably higher for shorter lactations, the analysis was performed using the "weights" option of the "relmatLmer"

function. For lactation periods of 305 days or more, a weight of 305 was taken, and for shorter durations, one equal to the actual number of lactation days. The Benjamini–Hochberg method (Benjamini, Hochberg, 1995) implemented in the "qvalue" R package (the "qvalue" function with "lambda=0" parameter) was applied to correct for multiple comparisons (Storey et al., 2023).

Results

Target fragments were successfully amplified for 136 (rs109039955), 130 (rs207681322) and 109 (rs110871255) animals. All three studied SNPs were found to be polymorphic in the studied sample (Table 1). The genotype distributions of rs207681322 and rs110871255 deviated significantly from Hardy–Weinberg equilibrium ($p = 0.0141$ and $p = 0.0039$, respectively), and that of rs109039955 did not. The LD between the studied SNPs was $r^2 = 0.098$, $D' = 0.853$ for rs109039955 and rs207681322; $r^2 = 0.003$, $D' = 0.061$ for rs109039955 and rs110871255; $r^2 = 0.001$, $D' = 0.088$ for rs207681322 and rs110871255.

Statistical tests revealed a total of six associations of two SNPs with three milk production traits from the 2nd to 4th lactations (Table 2). No associations of all three SNPs were detected for the first lactation. For the second lactation, one SNP (rs207681322) was associated with milk yield ($q = 0.043$). For the third lactation period, one SNP (rs110871255) was associated with fat yield ($q = 0.0135$) and another (rs207681322) with milk ($q = 2.54E-04$) and protein ($q = 0.021$) yields. For the fourth period, one SNP (rs110871255) was associated with fat yield ($q = 0.0348$) and another (rs207681322) with milk yield ($q = 0.021$).

Therefore, rs110871255 was associated with fat yield during the third and fourth lactations, and rs207681322 with milk yield in the second, third and fourth lactations as well as with protein yield in the third lactation.

Discussion

In the studied Yaroslavl cows, three SNPs (rs110871255, rs207681322 and rs109039955) with a rare allele frequency of 0.042–0.261 were identified. For the rs207681322 and rs110871255 loci, genotype distributions differed significantly

Table 1. Characteristics of the genotyped SNPs

SNP #	Position (ARS-UCD1.2)	Localization	Genotype	Genotype frequency	Allele	Allele frequency
rs110871255 (p.Met101Thr)	chr11:86127780	Exon 5	GG	0.725	G	0.826
			GA	0.202	A	0.174
			AA	0.073		
rs207681322 (p.Pro395Ser)	chr11:86116295	Exon 9	GG	0.931	G	0.958
			GA	0.054	A	0.042
			AA	0.015		
rs109039955	chr11:86082101	3'-UTR	GG	0.537	G	0.739
			GA	0.404	A	0.261
			AA	0.059		

Table 2. Associations of the studied SNPs with milk productivity traits from 1st to 4th lactations and their corresponding phenotypic values (mean ± standard deviation) for different genotypes

SNP#	Lactation	Genotype (Number of animals*)	Milk yield, kg	Fat yield, kg	Protein yield, kg	Fat %	Protein %
rs110871255	1	GG (78)	4698 ± 838	218.7 ± 45.5	148.5 ± 26.5	4.66 ± 0.56	3.17 ± 0.20
		GA (21)	4936 ± 818	225.4 ± 43.1	155.9 ± 24.4	4.57 ± 0.47	3.17 ± 0.23
		AA (7–8)	4441 ± 607	196.2 ± 34.9	142.5 ± 13.3	4.54 ± 0.69	3.30 ± 0.18
		q-value	0.7893	0.9428	0.6785	0.5054	0.3326
	2	GG (69)	5016 ± 692	230.9 ± 33.7	160.2 ± 22.0	4.63 ± 0.52	3.20 ± 0.22
		GA (20)	5341 ± 1011	245.9 ± 50.1	176.6 ± 35.8	4.62 ± 0.53	3.30 ± 0.24
		AA (7)	5237 ± 851	247.3 ± 56.7	166.2 ± 28.9	4.69 ± 0.53	3.18 ± 0.22
		q-value	0.3782	0.1826	0.3406	0.4299	0.7893
	3	GG (56)	5367 ± 617	242.9 ± 35.3	168.2 ± 24.7	4.54 ± 0.56	3.13 ± 0.22
		GA (15)	5925 ± 1209	275.1 ± 38.4	196.0 ± 41.8	4.72 ± 0.58	3.31 ± 0.22
		AA (6)	5903 ± 843	281.8 ± 51.5	191.3 ± 20.9	4.79 ± 0.71	3.26 ± 0.23
		q-value	0.0825	0.0135	0.072	0.2986	0.1826
	4	GG (41)	5485 ± 861	251.1 ± 42.8	179.2 ± 28.4	4.61 ± 0.69	3.27 ± 0.18
		GA (12)	6222 ± 885	291.9 ± 30.7	203.1 ± 28.4	4.74 ± 0.54	3.27 ± 0.21
		AA (4)	5836 ± 419	269.9 ± 41.7	186.1 ± 23.6	4.61 ± 0.49	3.18 ± 0.20
		q-value	0.1826	0.0348	0.2986	0.5931	0.4299
rs207681322	1	GG (117–118)	4784 ± 873	220.9 ± 46.3	150.8 ± 25.9	4.63 ± 0.52	3.17 ± 0.21
		GA (7)	5034 ± 801	211.2 ± 48.8	153.7 ± 18.2	4.16 ± 0.41	3.07 ± 0.15
		AA (2)	5502 ± 503	239.6 ± 4.50	164.4 ± 12.0	4.37 ± 0.32	2.99 ± 0.06
		q-value	0.211	0.6785	0.5911	0.3264	0.1728
	2	GG (106)	5106 ± 791	232.8 ± 37.1	163.2 ± 25.4	4.58 ± 0.50	3.20 ± 0.23
		GA (5)	6148 ± 739	278.2 ± 55.2	201.7 ± 29.3	4.50 ± 0.41	3.28 ± 0.18
		AA (2)	5655 ± 1059	257.0 ± 73.6	166.9 ± 32.1	4.51 ± 0.46	2.95 ± 0.01
		q-value	0.043	0.12	0.1343	0.7904	0.3326
	3	GG (85)	5458 ± 835	245.5 ± 39.7	174.0 ± 29.1	4.53 ± 0.62	3.19 ± 0.23
		GA (4)	7116 ± 1560	289.7 ± 50.4	224.1 ± 61.0	4.11 ± 0.39	3.13 ± 0.16
		AA (1)	8195	303.6	238.7	3.70	2.91
		q-value	2.54E-04	0.1826	0.021	0.104	0.3406
	4	GG (55)	5587 ± 883	258.7 ± 44.1	182.1 ± 29.9	4.66 ± 0.64	3.26 ± 0.19
		GA (3)	6320 ± 303	277.0 ± 44.6	199.6 ± 19.2	4.37 ± 0.57	3.15 ± 0.17
		AA (1)	8836	304.3	264.2	3.44	2.99
		q-value	0.021	0.6399	0.12	0.111	0.1826
rs109039955	1	GG (73)	4732 ± 841	218.5 ± 48.1	149.9 ± 25.1	4.61 ± 0.51	3.18 ± 0.20
		GA (52)	4812 ± 855	217.5 ± 39.7	150.9 ± 26.3	4.54 ± 0.51	3.15 ± 0.21
		AA (8)	4883 ± 697	232.2 ± 29.4	152.4 ± 16.2	4.80 ± 0.74	3.14 ± 0.21
		q-value	0.6087	0.6785	0.7977	0.7782	0.3264
	2	GG (66)	5067 ± 713	230.0 ± 36.5	163.3 ± 24.9	4.55 ± 0.47	3.23 ± 0.24
		GA (44)	5260 ± 856	238.1 ± 39.9	166.9 ± 28.5	4.55 ± 0.50	3.18 ± 0.22
		AA (7)	5475 ± 838	260.1 ± 46.9	177.0 ± 30.0	4.78 ± 0.75	3.24 ± 0.25
		q-value	0.1826	0.1728	0.3148	0.7904	0.624
	3	GG (51)	5415 ± 697	243.2 ± 37.5	172.7 ± 26.8	4.50 ± 0.53	3.19 ± 0.24
		GA (35)	5541 ± 1053	245.6 ± 41.0	175.1 ± 37.0	4.48 ± 0.60	3.16 ± 0.21
		AA (6)	6101 ± 1322	278.2 ± 45.1	192.3 ± 34.1	4.65 ± 0.84	3.18 ± 0.24
		q-value	0.23	0.1849	0.4253	0.9213	0.6087
	4	GG (35)	5560 ± 974	251.8 ± 41.7	181.7 ± 32.3	4.56 ± 0.55	3.27 ± 0.20
		GA (20)	5815 ± 652	268.3 ± 46.7	186.7 ± 22.8	4.61 ± 0.67	3.21 ± 0.18
		AA (6)	5668 ± 1901	266.3 ± 61.1	181.1 ± 56.7	4.88 ± 1.07	3.21 ± 0.21
		q-value	0.7114	0.5911	0.9722	0.7893	0.3148

Note. The statistically significant associations ($q < 0.05$) are indicated in bold.

* Only animals having respective phenotypic data were accounted for. The markedly lower number of animals genotyped for rs110871255 was due to the limited amount of DNA available for analysis.

from those expected by Hardy–Weinberg equilibrium due to an excess of rare homozygotes, which may be due to inbreeding, gene drift, or selection in farm animal populations (Hedrick, 2005). Such deviations are often observed in studies of microsatellite DNA markers or SNPs in different cattle breeds (Melka, Schenkel, 2012; Madilindi et al., 2020; Ocampo et al., 2021). It is noteworthy that tests for Hardy–Weinberg equilibrium deviation are used to verify random mating in populations, while tests for deviations from expected homozygote frequency are performed to estimate inbreeding coefficients (Haldane, 1984; Robertson, Hill, 1984). The LD between the studied SNPs in Yaroslavl cows differs significantly from that described previously for the same SNPs in Chinese Holstein cows (Han B. et al., 2019), but these patterns are known to vary significantly between cattle breeds (Porto-Neto et al., 2014).

For two SNPs, we found significant associations with such traits as milk, fat and protein yields across multiple lactations, e. g., allele A of rs207681322 had a positive effect on the cows' milk and protein yields. Previously, these associations were found in the first or second lactation in Chinese Holsteins (Han B. et al., 2019). In our study, the A allele of rs110871255 positively affected fat yield. However, in Holstein cows, the same allele was associated with increased fat percentage during the second lactation period (Han B. et al., 2019).

Unlike Han B. et al. (2019), the associations we found were in subsequent lactations rather than in the first one, which may be due to both genetic (different breeds) and environmental (different breeding conditions) factors. It should be emphasized that there is still no consensus on the effect hereditary factors have on milk productivity during lactation. While some authors believe that the heritability of these traits in the Holstein breed for the first lactation is higher than for subsequent ones (Dimov et al., 1995; Yamazaki et al., 2016; Lee et al., 2020), the latest large-scale study of nearly 3.5 million records from over 1 million Holstein cows found increased heritability of milk, fat and protein yields specifically in lactations 3 and 4 compared to lactations 1 and 2 (Williams et al., 2022), which agrees well with our results. Milk yield heritability for individual test days has been significantly higher during the third lactation compared to the first lactation in holsteinized native cattle from Thailand (Buaban et al., 2020) and Sahiwal cows from Kenya (Ilatsia et al., 2007).

The associations we found correspond well with the data of other authors on the effect *LPIN1* has on lactation, lipid biosynthesis and general metabolic levels. The gene's mRNA expression has significantly increased during lactation in the mammary gland of humans (Mohammad, Haymond, 2013), pigs (Lv et al., 2015) and mice (Han L.Q. et al., 2010). SNP in the *LPIN1* gene is associated with the percentage of visceral and intramuscular fat in pigs (He et al., 2009). The gene's mutation leading to the expression of a truncated protein causes lipodystrophy in rats (Mul et al., 2011). The FLD mice with mutated *LPIN1* have a phenotype similar to that of hereditary lipodystrophy in humans characterized by subcutaneous fat loss, hepatic steatosis, insulin resistance, etc. (Péterfy et al., 2001). Conversely, *LPIN1* overexpression in adipose tissues or skeletal muscles causes obesity in transgenic mice (Phan, Reue, 2005). Moreover, while *LPIN1* expression in adipose tissue stimulates adipocytes to accumulate fat, in muscle tis-

sue its expression affects the whole body's energy expenditure (temperature; food and oxygen consumption). Suppression of *LPIN1* expression in muscle cells *in vitro* has led to insulin resistance (Huang et al., 2017). At the same time, a negative correlation was observed between *LPIN1* mRNA levels in adipose tissue and blood glucose/insulin concentrations in humans and mice (Suviolahti et al., 2006). The same authors found the SNPs in the *LPIN1* gene are associated with blood insulin levels in dyslipidemia families (Suviolahti et al., 2006). In their opinion, *LPIN1* plays an essential role in glucose homeostasis and its genetic variants affect metabolism traits. Indeed, genetic variations of the gene are associated with the development of some metabolic syndromes in humans (Brahe et al., 2013; Zhang et al., 2013).

Conclusion

Our study has found the rs207681322 and rs110871255 SNPs in the *LPIN1* gene have statistically significant associations with some milk productivity traits during several lactations in the Yaroslavl cows to be previously found by other authors in the Holstein breed. The results obtained can be used in marker-assisted and genomic selection for dairy cattle breeding.

References

- Abdelmanova A.S., Kharzinova V.R., Volkova V.V., Mishina A.I., Dotssev A.V., Sermyagin A.A., Boronetskaya O.I., Petrikeeva L.V., Chinarov R.Y., Brem G., Zinovieva N.A. Genetic diversity of historical and modern populations of Russian cattle breeds revealed by microsatellite analysis. *Genes (Basel)*. 2020;11(8):940. DOI 10.3390/genes11080940
- Ahmad S.M., Bhat B., Bhat S.A., Yaseen M., Mir S., Raza M., Iqbal M.A., Shah R.A., Ganai N.A. SNPs in mammary gland epithelial cells unraveling potential difference in milk production between Jersey and Kashmiri cattle using RNA sequencing. *Front. Genet.* 2021;12:666015. DOI 10.3389/fgene.2021.666015
- Barroso E., Rodríguez-Calvo R., Serrano-Marco L., Astudillo A.M., Balsile J., Palomer X., Vázquez-Carrera M. The PPAR β/δ activator GW501516 prevents the down-regulation of AMPK caused by a high-fat diet in liver and amplifies the PGC-1 α -Lipin 1-PPAR α pathway leading to increased fatty acid oxidation. *Endocrinology*. 2011;152(5):1848-1859. DOI 10.1210/en.2010-1468
- Bekele R., Taye M., Abebe G., Meseret S. Genomic regions and candidate genes associated with milk production traits in Holstein and its crossbred cattle: a review. *Int. J. Genomics*. 2023;2023:8497453. DOI 10.1155/2023/8497453
- Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 1995;57(1):289-300. DOI 10.1111/j.2517-6161.1995.tb02031
- Bionaz M., Looor J.J. *ACSL1*, *AGPAT6*, *FABP3*, *LPIN1*, and *SLC27A6* are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. *J. Nutr.* 2008;138(6):1019-1024. DOI 10.1093/jn/138.6.1019
- Brahe L.K., Ångquist L., Larsen L.H., Vimaleswaran K.S., Hager J., Viguerie N., Loos R.J., Handjieva-Darlenska T., Jebb S.A., Hlavaty P., Larsen T.M., Martinez J.A., Papadaki A., Pfeiffer A.F., van Baak M.A., Sørensen T.I., Holst C., Langin D., Astrup A., Saris W.H. Influence of SNPs in nutrient-sensitive candidate genes and gene-diet interactions on blood lipids: the DiOGenes study. *Br. J. Nutr.* 2013;110(5):790-796. DOI 10.1017/S0007114512006058
- Buaban S., Puangdee S., Duangjinda M., Boonkum W. Estimation of genetic parameters and trends for production traits of dairy cattle in Thailand using a multiple-trait multiple-lactation test day model.

- Asian-Australas. J. Anim. Sci.* 2020;33(9):1387-1399. DOI 10.5713/ajas.19.0141
- Cecchinato A., Ribeca C., Chessa S., Cipolat-Gotet C., Maretto F., Casellas J., Bittante G. Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. *Animal*. 2014;8(7):1062-1070. DOI 10.1017/S1751731114001098
- Chen S.Y., Gloria L.S., Pedrosa V.B., Doucette J., Boerman J.P., Brito L.F. Unravelling the genomic background of resilience based on variability in milk yield and milk production levels in North American Holstein cattle through GWAS and Mendelian randomization analyses. *J. Dairy Sci.* 2023;107:1035-1053. DOI 10.3168/jds.2023-23650
- Chen Y., Rui B.B., Tang L.Y., Hu C.M. Lipin family proteins – key regulators in lipid metabolism. *Ann. Nutr. Metab.* 2015;66(1):10-18. DOI 10.1159/000368661
- Csaki L.S., Dwyer J.R., Fong L.G., Tontonoz P., Young S.G., Reue K. Lipins, lipinopathies, and the modulation of cellular lipid storage and signaling. *Prog. Lipid Res.* 2013;52(3):305-316. DOI 10.1016/j.plipres.2013.04.001
- Dimov G., Albuquerque L.G., Keown J.F., Van Vleck L.D., Norman H.D. Variance of interaction effects of sire and herd for yield traits of Holsteins in California, New York, and Pennsylvania with an animal model. *J. Dairy Sci.* 1995;78(4):939-946. DOI 10.3168/jds.S0022-0302(95)76709-1
- Dmitriev N.G. Breed Cattle by Countries of the World. Leningrad: Kolos Publ., 1978 (in Russian)
- Dmitriev N.G., Ernst L.K. (Eds.) Animal Genetics Resources of the USSR. Rome: Food and Agriculture Organization of the United Nations, 1989
- Du X., Zhou H., Liu X., Li Y., Hickford J.G.H. Sequence variation in the bovine lipin-1 gene (*LPIN1*) and its association with milk fat and protein contents in New Zealand Holstein-Friesian × Jersey (HF × J)-cross dairy cows. *Animals (Basel)*. 2021;11(11):3223. DOI 10.3390/ani11113223
- Dunin I.M., Dankvert A.G. (Eds.) Breeds and Types of Farm Animals in the Russian Federation. Moscow: All-Russia Research Institute of Animal Breeding, 2013 (in Russian)
- Fang L., Cai W., Liu S., Canela-Xandri O., Gao Y., Jiang J., Rawlik K., Li B., Schroeder S.G., Rosen B.D., Li C.J., Sonstegard T.S., Alexander L.J., Van Tassell C.P., Van Raden P.M., Cole J.B., Yu Y., Zhang S., Tenesa A., Ma L., Liu G.E. Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res.* 2020;30(5):790-801. DOI 10.1101/gr.250704.119
- Gutierrez-Reinoso M.A., Aponte P.M., Garcia-Herreros M. Genomic analysis, progress and future perspectives in dairy cattle selection: a review. *Animals (Basel)*. 2021;11(3):599. DOI 10.3390/ani11030599
- Haldane J.B.S. An exact test for randomness of mating. *J. Genet.* 1954; 52:631-635. DOI 10.1007/BF02985085
- Han B., Yuan Y., Liang R., Li Y., Liu L., Sun D. Genetic effects of *LPIN1* polymorphisms on milk production traits in dairy cattle. *Genes (Basel)*. 2019;10(4):265. DOI 10.3390/genes10040265
- Han L.Q., Li H.J., Wang Y.Y., Zhu H.S., Wang L.F., Guo Y.J., Lu W.F., Wang Y.L., Yang G.Y. mRNA abundance and expression of SLC27A, ACC, SCD, FADS, LPIN, INSIG, and PPARGC1 gene isoforms in mouse mammary glands during the lactation cycle. *Genet. Mol. Res.* 2010;9(2):1250-1257. DOI 10.4238/vol9-2gmr814
- He X.P., Xu X.W., Zhao S.H., Fan B., Yu M., Zhu M.J., Li C.C., Peng Z.Z., Liu B. Investigation of *Lpin1* as a candidate gene for fat deposition in pigs. *Mol. Biol. Rep.* 2009;36(5):1175-1180. DOI 10.1007/s11033-008-9294-4
- Hedrick P.W. Genetics of Populations. Jones & Bartlett Learning, 2005
- Hillers J.K., Williams G.F. "Shook" factors and lactation milk yield. *Extension Bull. (Wash. State Univ.)*. 1981;EB0779
- Huang S., Huang S., Wang X., Zhang Q., Liu J., Leng Y. Downregulation of lipin-1 induces insulin resistance by increasing intracellular ceramide accumulation in C2C12 myotubes. *Int. J. Biol. Sci.* 2017;13(1):1-12. DOI 10.7150/ijbs.17149
- Ilatsia E.D., Muasya T.K., Muhuyi W.B., Kahi A.K. Genetic and phenotypic parameters for test day milk yield of Sahiwal cattle in the semi-arid tropics. *Animal*. 2007;1(2):185-192. DOI 10.1017/S175173110739263X
- Iso-Touru T., Tapio M., Vilkki J., Kiseleva T., Ammosov I., Ivanova Z., Popov R., Ozerov M., Kantanen J. Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim. Genet.* 2016;47(6):647-657. DOI 10.1111/age.12473
- Kadegowda A.K., Bionaz M., Piperova L.S., Erdman R.A., Looor J.J. Peroxisome proliferator-activated receptor-gamma activation and long-chain fatty acids alter lipogenic gene networks in bovine mammary epithelial cells to various extents. *J. Dairy Sci.* 2009;92(9):4276-4289. DOI 10.3168/jds.2008-1932
- Khatkar M.S., Thomson P.C., Tammen I., Raadsma H.W. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol.* 2004;36(2):163-190. DOI 10.1186/1297-9686-36-2-163
- Kim H.E., Bae E., Jeong D.Y., Kim M.J., Jin W.J., Park S.W., Han G.S., Carman G.M., Koh E., Kim K.S. Lipin1 regulates PPARγ transcriptional activity. *Biochem. J.* 2013;453(1):49-60. DOI 10.1042/BJ20121598
- Lee Y.M., Dang C.G., Alam M.Z., Kim Y.S., Cho K.H., Park K.D., Kim J.J. The effectiveness of genomic selection for milk production traits of Holstein dairy cattle. *Asian-Australas. J. Anim. Sci.* 2020; 33(3):382-389. DOI 10.5713/ajas.19.0546
- Li Q., Liang R., Li Y., Gao Y., Li Q., Sun D., Li J. Identification of candidate genes for milk production traits by RNA sequencing on bovine liver at different lactation stages. *BMC Genet.* 2020;21(1):72. DOI 10.1186/s12863-020-00882-y
- Lopdell T.J. Using QTL to identify genes and pathways underlying the regulation and production of milk components in cattle. *Animals (Basel)*. 2023;13(5):911. DOI 10.3390/ani13050911
- Lu G., Moriyama E.N. Vector NTI, a balanced all-in-one sequence analysis suite. *Brief. Bioinform.* 2004;5(4):378-388. DOI 10.1093/bib/5.4.378
- Lv Y., Guan W., Qiao H., Wang C., Chen F., Zhang Y., Liao Z. Veterinary Medicine and Omics (Veterinomics): metabolic transition of milk triacylglycerol synthesis in sows from late pregnancy to lactation. *OMICS*. 2015;19(10):602-616. DOI 10.1089/omi.2015.0102
- Madilindi M.A., Banga C.B., Bhebhe E., Sanarana Y.P., Nxumalo K.S., Taela M.G., Magagula B.S., Mapholi N.O. Genetic diversity and relationships among three Southern African Nguni cattle populations. *Trop. Anim. Health Prod.* 2020;52(2):753-762. DOI 10.1007/s11250-019-02066-y
- Melka M.G., Schenkel F.S. Analysis of genetic diversity in Brown Swiss, Jersey and Holstein populations using genome-wide single nucleotide polymorphism markers. *BMC Res. Notes*. 2012;5:161. DOI 10.1186/1756-0500-5-161
- Mohammad M.A., Haymond M.W. Regulation of lipid synthesis genes and milk fat production in human mammary epithelial cells during secretory activation. *Am. J. Physiol. Endocrinol. Metab.* 2013; 305(6):E700-E716. DOI 10.1152/ajpendo.00052.2013
- Mul J.D., Nadra K., Jagalur N.B., Nijman I.J., Toonen P.W., Médard J.J., Grès S., de Bruin A., Han G.S., Brouwers J.F., Carman G.M., Saulnier-Blache J.S., Meijer D., Chrast R., Cuppen E. A hypomorphic mutation in *Lpin1* induces progressively improving neuropathy and lipodystrophy in the rat. *J. Biol. Chem.* 2011;286(30):26781-26793. DOI 10.1074/jbc.M110.197947
- Nayak S.S., Panigrahi M., Rajawat D., Ghildiyal K., Sharma A., Parida S., Bhushan B., Mishra B.P., Dutt T. Comprehensive selection signature analyses in dairy cattle exploiting purebred and crossbred

- genomic data. *Mamm. Genome*. 2023;34(4):615-631. DOI 10.1007/s00335-023-10021-4
- Ocampo R.J., Martínez J.F., Martínez R. Assessment of genetic diversity and population structure of Colombian Creole cattle using microsatellites. *Trop. Anim. Health Prod.* 2021;53(1):122. DOI 10.1007/s11250-021-02563-z
- Pegolo S., Cecchinato A., Mele M., Conte G., Schiavon S., Bittante G. Effects of candidate gene polymorphisms on the detailed fatty acids profile determined by gas chromatography in bovine milk. *J. Dairy Sci.* 2016;99(6):4558-4573. DOI 10.3168/jds.2015-10420
- Persichilli C., Senczuk G., Mastrangelo S., Marusi M., van Kaam J.T., Finocchiaro R., Di Civita M., Cassandro M., Pilla F. Exploring genome-wide differentiation and signatures of selection in Italian and North American Holstein populations. *J. Dairy Sci.* 2023;106(8):5537-5553. DOI 10.3168/jds.2022-22159
- Péterfy M., Phan J., Xu P., Reue K. Lipodystrophy in the *fld* mouse results from mutation of a new gene encoding a nuclear protein, lipin. *Nat. Genet.* 2001;27(1):121-124. DOI 10.1038/83685
- Peterson T.R., Sengupta S.S., Harris T.E., Carmack A.E., Kang S.A., Balderas E., Guertin D.A., Madden K.L., Carpenter A.E., Finck B.N., Sabatini D.M. mTOR complex 1 regulates lipin 1 localization to control the SREBP pathway. *Cell*. 2011;146(3):408-420. DOI 10.1016/j.cell.2011.06.034
- Phan J., Reue K. Lipin, a lipodystrophy and obesity gene. *Cell Metab.* 2005;1(1):73-83. DOI 10.1016/j.cmet.2004.12.002
- Porto-Neto L.R., Kijas J.W., Reverter A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet. Sel. Evol.* 2014;46(1):22. DOI 10.1186/1297-9686-46-22
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J., Sham P.C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007;81(3):559-575. DOI 10.1086/519795
- Rajawat D., Panigrahi M., Kumar H., Nayak S.S., Parida S., Bhushan B., Gaur G.K., Dutt T., Mishra B.P. Identification of important genomic footprints using eight different selection signature statistics in domestic cattle breeds. *Gene*. 2022;816:146165. DOI 10.1016/j.gene.2021.146165
- Reue K., Zhang P. The lipin protein family: dual roles in lipid biosynthesis and gene expression. *FEBS Lett.* 2008;582(1):90-96. DOI 10.1016/j.febslet.2007.11.014
- Robertson A., Hill W.G. Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics*. 1984;107(4):703-718. DOI 10.1093/genetics/107.4.703
- Ruvinskiy D., Igoshin A., Yurchenko A., Ilina A.V., Larkin D.M. Resequencing the Yaroslavl cattle genomes reveals signatures of selection and a rare haplotype on BTA28 likely to be related to breed phenotypes. *Anim. Genet.* 2022;53(5):680-684. DOI 10.1111/age.13230
- Sambrook J., Russell D.W. *The Condensed Protocols from Molecular Cloning: a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press, 2006
- Saydakova S.S., Morozova K.N., Kiseleva E.V. Lipin family proteins: structure, functions, and related diseases. *Cell Tissue Biol.* 2021;15(4):317-325. DOI 10.1134/S1990519X21040076
- Sermyagin A.A., Dotsev A.V., Gladyr E.A., Traspov A.A., Denisikova T.E., Kostyunina O.V., Reyher H., Wimmers K., Barbato M., Paronyan I.A., Plemnyashov K.V., Sölkner J., Popov R.G., Brem G., Zinovieva N.A. Whole-genome SNP analysis elucidates the genetic structure of Russian cattle and its relationship with Eurasian taurine breeds. *Genet. Sel. Evol.* 2018;50(1):37. DOI 10.1186/s12711-018-0408-8
- Shichkin G.I., Tyapugin E.E., Dunin I.M., Gerasimova E.V., Kozlova N.A., Myshkina M.S., Semenova N.V., Dunin M.I., Tyapugin S.E. The state of dairy cattle breeding in the Russian Federation. In: *Yearbook on Breeding Work in Dairy Cattle Breeding in Farms of the Russian Federation*. Moscow: All-Russia Research Institute of Animal Breeding, 2023;3-20 (in Russian)
- Silpa M.V., König S., Sejian V., Malik P.K., Nair M.R.R., Fonseca V.F.C., Maia A.S.C., Bhatta R. Climate-resilient dairy cattle production: applications of genomic tools and statistical models. *Front. Vet. Sci.* 2021;8:625189. DOI 10.3389/fvets.2021.625189
- Singh A., Malla W.A., Kumar A., Jain A., Thakur M.S., Khare V., Tiwari S.P. Review: genetic background of milk fatty acid synthesis in bovines. *Trop. Anim. Health Prod.* 2023;55(5):328. DOI 10.1007/s11250-023-03754-6
- Siniossoglou S. Phospholipid metabolism and nuclear function: roles of the lipin family of phosphatidic acid phosphatases. *Biochim. Biophys. Acta*. 2013;1831(3):575-581. DOI 10.1016/j.bbali.2012.09.014
- Sinnwell J.P., Therneau T.M., Schaid D.J. The kinship2 R package for pedigree data. *Hum. Hered.* 2014;78(2):91-93. DOI 10.1159/000363105
- Stolpovsky Yu.A., Gosteva E.R., Solodneva E.V. Genetic and Selection Aspects of the History of the Development of Cattle Breeding on the Territory of Russia. Moscow: Akvarel' Publ., 2022 (in Russian)
- Storey J.D., Bass A.J., Dabney A., Robinson D. qvalue: Q-value estimation for false discovery rate control. R Package version 2.34.0. 2023. DOI 10.18129/B9.bioc.qvalue. (<https://bioconductor.org/packages/qvalue>)
- Suviolahti E., Reue K., Cantor R.M., Phan J., Gentile M., Naukkarienen J., Soro-Paavonen A., Oksanen L., Kaprio J., Rissanen A., Salomaa V., Kontula K., Taskinen M.R., Pajukanta P., Peltonen L. Cross-species analyses implicate Lipin 1 involvement in human glucose metabolism. *Hum. Mol. Genet.* 2006;15(3):377-386. DOI 10.1093/hmg/ddi448
- Teng J., Wang D., Zhao C., Zhang X., Chen Z., Liu J., Sun D., Tang H., Wang W., Li J., Mei C., Yang Z., Ning C., Zhang Q. Longitudinal genome-wide association studies of milk production traits in Holstein cattle using whole-genome sequence data imputed from medium-density chip data. *J. Dairy Sci.* 2023;106(4):2535-2550. DOI 10.3168/jds.2022-22277
- Thering B.J., Graugnard D.E., Piantoni P., Looor J.J. Adipose tissue lipogenic gene networks due to lipid feeding and milk fat depression in lactating cows. *J. Dairy Sci.* 2009;92(9):4290-4300. DOI 10.3168/jds.2008-2000
- Weller J.I., Ron M. Invited review: quantitative trait nucleotide determination in the era of genomic selection. *J. Dairy Sci.* 2011;94(3):1082-1090. DOI 10.3168/jds.2010-3793
- Weller J.I., Ezra E., Ron M. Invited review: a perspective on the future of genomic selection in dairy cattle. *J. Dairy Sci.* 2017;100(11):8633-8644. DOI 10.3168/jds.2017-12879
- Wiggins G.R., Van Vleck L.D. Extending partial lactation milk and fat records with a function of last-sample production. *J. Dairy Sci.* 1979;62(2):316-325. DOI 10.3168/jds.S0022-0302(79)83242-7
- Williams M., Sleator R.D., Murphy C.P., McCarthy J., Berry D.P. Exploiting genetic variability in the trajectory of lactation yield and somatic cell score with each progressing parity. *J. Dairy Sci.* 2022;105(4):3341-3354. DOI 10.3168/jds.2021-21306
- Yamazaki T., Hagiya K., Takeda H., Osawa T., Yamaguchi S., Nagamine Y. Effects of stage of pregnancy on variance components, daily milk yields and 305-day milk yield in Holstein cows, as estimated by using a test-day model. *Animal*. 2016;10(8):1263-1270. DOI 10.1017/S1751731116000185
- Ye J., Coulouris G., Zaretskaya I., Cutcutache I., Rozen S., Madden T.L. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13:134. DOI 10.1186/1471-2105-13-134
- Yurchenko A.A., Daetwyler H.D., Yudin N., Schnabel R.D., Vander Jagt C.J., Soloshenko V., Lhasaranov B., Popov R., Taylor J.F., Lar-

- kin D.M. Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Sci. Rep.* 2018a;8(1):12984. DOI 10.1038/s41598-018-31304-w
- Yurchenko A., Yudin N., Aitnazarov R., Plyusnina A., Brukhin V., Soloshenko V., Lhasaranov B., Popov R., Paronyan I.A., Plemyshev K.V., Larkin D.M. Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds. *Heredity (Edinb)*. 2018b;120(2):125-137. DOI 10.1038/s41437-017-0024-3
- Zhang R., Jiang F., Hu C., Yu W., Wang J., Wang C., Ma X., Tang S., Bao Y., Xiang K., Jia W. Genetic variants of *LPIN1* indicate an association with Type 2 diabetes mellitus in a Chinese population. *Diabet. Med.* 2013;30(1):118-122. DOI 10.1111/j.1464-5491.2012.03758.x
- Zinovieva N.A., Dotsev A.V., Sermyagin A.A., Deniskova T.E., Abdelmanova A.S., Kharzinova V.R., Sölkner J., Reyer H., Wimmers K., Brem G. Selection signatures in two oldest Russian native cattle breeds revealed using high-density single nucleotide polymorphism analysis. *PLoS One.* 2020;15(11):e0242200. DOI 10.1371/journal.pone.0242200
- Ziyatdinov A., Vázquez-Santiago M., Brunel H., Martínez-Perez A., Aschard H., Soria J.M. lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics.* 2018;19(1):68. DOI 10.1186/s12859-018-2057-x

ORCID

A.V. Igoshin orcid.org/0000-0002-1896-0072
R.B. Aitnazarov orcid.org/0000-0001-6929-7648
A.V. Ilina orcid.org/0000-0002-5046-1812
D.M. Larkin orcid.org/0000-0001-7859-6201
N.S. Yudin orcid.org/0000-0002-1947-5554

Acknowledgements. This work was supported by State Budgeted Project No. FWNR-2022-0039.

Conflict of interest. The authors declare no conflict of interest.

Received December 1, 2023. Revised January 15, 2024. Accepted January 15, 2024.

Прием статей через электронную редакцию на сайте <http://vavilov.elpub.ru/index.php/jour>
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции»/“Vavilov Journal of Genetics and Breeding”
до 2011 г. выходил под названием «Информационный вестник ВОГиС»/
“The Herald of Vavilov Society for Geneticists and Breeding Scientists”.

Сетевое издание «Вавиловский журнал генетики и селекции» – реестровая запись СМИ
Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий,
в которых должны быть опубликованы основные результаты диссертаций на соискание ученой
степени кандидата наук, на соискание ученой степени доктора наук, Russian Science Citation Index
на платформе Web of Science, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC,
Scopus, PubMed Central, DOAJ, ROAD, Ulrich’s Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:

русскоязычная версия – на сайте ИЦиГ СО РАН, <https://vavilovj-icg.ru/>
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440

англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, <https://www.ncbi.nlm.nih.gov/pmc/journals/3805/>

При перепечатке материалов ссылка обязательна.

✉ email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.

Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.

Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.

Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.

Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.

Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.

Дата публикации 29.02.2024. Формат 60 × 84 1/8. Уч.-изд. л. 16.3.